8-2018

Biological Condition and Stressors of BLM Wadeable Streams in Northeastern California and Northwestern Nevada

Nicole Cappuccio
Utah State University

BIOLOGICAL CONDITION AND STRESSORS OF BLM WADEABLE

STREAMS IN NORTHEASTERN CALIFORNIA AND

NORTHWESTERN NEVADA

by

Nicole Cappuccio

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Ecology

Approved:

Scott W. Miller, Ph.D. Major Professor	Charles P. Hawkins, Ph.D. Committee Member
Sarah Null, Ph.D. Committee Member	Mark R. McLellan, Ph.D. Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2018

ABSTRACT

Biological Condition and Stressors of BLM Wadeable Streams in Northeastern California and Northwestern Nevada

by

Nicole Cappuccio, Master of Science

Utah State University, 2018

Major Professor: Dr. Scott W. Miller
Department: Watershed Sciences

Public land management agencies spend millions of dollars annually to monitor aquatic resource condition and trend, but few implement standard, agency-wide protocols, which results in data that cannot be compared through space or time or among agencies. To address this challenge, the Bureau of Land Management (BLM) developed the National Aquatic Monitoring Framework (NAMF). As one of the first applications of the NAMF, the objectives of this study were to determine the biological condition of lotic systems, the extent of stressors and their impact to biological condition, and the anthropogenic sources of stressors. I selected a spatially balanced random sample of sites on BLM land in Northeast California and Northwest Nevada. I used a macroinvertebrate multimetric index (MMI) and found 45.3% of stream km within the study area have degraded biological condition. Of the chemical and physical stressors, total nitrogen concentration (TN), total phosphorus concentration (TP), and altered canopy cover were most pervasive. I found 68% of stream km have excessive total nitrogen, 43% have canopy cover below expected conditions, and 37% have excessive total phosphorus. I used random forest models to explain 28% of the variability in MMI scores with watershed area, riparian complexity, TN, intermittent stream density, and TP (in order of importance). This result indicates that TN, TP, and riparian complexity were the most biologically relevant stressors. I then used random forest models to identify land uses associated with biologically relevant and most geographically extensive stressors. I included natural predictors in these models to identify potential interactions between land uses and natural variables. Excess TN and TP were associated with livestock grazing intensity and duration. However, riparian complexity and canopy cover were only associated with natural predictors indicating much of the spatial variability in riparian values was naturally occurring. For public land management agencies, identifying priority stressors (most pervasive or most biologically harmful) and their likely sources is critical to effective and efficient adaptive management. This study provides an example of the quantitative data and analytical framework needed to assess the overall efficacy of management actions, aid adaptive management decisions, and ultimately ensure compliance with federal regulations.

(75 pages)

PUBLIC ABSTRACT

Biological Condition and Stressors of BLM Wadeable Streams in Northeastern California and Northwestern Nevada

Nicole Cappuccio

Taxpayer dollars can be used more efficiently by land management agencies to monitor streams if agency-wide monitoring protocols are adopted. To address this issue, the Bureau of Land Management (BLM) developed the National Aquatic Monitoring Framework (NAMF) to implement standardized assessments of stream condition and trend in the Western United States. As one of the first applications of the NAMF I sought to develop and apply an analytical framework to determine the biological condition of streams, extent of instream stressors and their impact on biological condition, and anthropogenic sources of stressors in Northeast California and Northwest Nevada over three years at a cost of $80,000. I measured biological, chemical, and physical attributes to determine the condition of streams at 70 spatially distributed random locations. I found 45% of BLM stream km in the study area have degraded biology, 68% have excessive total nitrogen (TN), 43% have canopy cover below expected conditions, and 37% have excessive total phosphorus (TP). Excessive TN and TP and degraded riparian complexity (RC) were most strongly related to degraded biological conditions. The occurrence of excess TN and TP was most associated with livestock grazing. RC was identified as a stressor, but was not associated with land uses. This study provides an example of the data and analytical approach needed to help the BLM adaptively manage streams and rivers in compliance with federal regulations while efficiently using taxpayer dollars.

ACKNOWLEDGMENTS

I would like to express my gratitude to the Bureau of Land Management (BLM) for funding this study and the opportunity to work on such an applied project in a unique and beautiful landscape. I am especially thankful for my advisor Scott Miller who helped me improve as a scientist, writer, and an individual while being a wonderful and patient mentor. Additionally I would like to thank my committee members Chuck Hawkins and Sarah Null for their great advice and support. I thank the BLM field office personnel from Eagle Lake, Surprise, and Alturas Field Offices for all of their help during my fieldwork, and overall support and encouragement. I am grateful to Tony Olsen and Phil Kaufmann for all of their help understanding and implementing the EPA methods and analyses. I am extremely thankful for the help and encouragement from my colleagues John Olsen, Jacob Vander Laan, Jennifer Courtwright, Robin Jones, all of my field technicians, and the Utah State University Bug Lab taxonomists and sorters. A special thanks to my late colleague Sarah Judson, who we sadly lost in 2015, for sharing such great knowledge, advice, and support. Lastly, thank you to all of my friends and family for the support you have offered me throughout my thesis process.

Nicole Cappuccio

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The objective of the Clean Water Act of 1972 (CWA) is “to restore and maintain the chemical, physical, and biological integrity of the Nation’s waters.” To achieve this objective, managers require accurate inventories of the type and location of aquatic resources, tools to quantify their condition and trend, and methods to identify the stressors associated with degraded conditions (Paulsen et al. 1998, Hawkins et al. 2008, Vander Laan et al. 2013). To date, regulatory agencies have developed programs to monitor the condition of lotic (i.e., streams and rivers) systems at the state, tribal, and national levels (e.g., California's Surface Water Ambient Monitoring Program and Environmental Protection Agency's [EPA] National Rivers and Streams Assessment [NRSA]). While these monitoring programs encompass the nation’s public lands, they frequently do not result in adequate sample sizes to accurately report on the condition and trend of freshwater resources found on public lands (e.g., Paulsen et al. 2008).

Public land management agencies spend millions of dollars annually to monitor aquatic resource condition and trend (ITFM 1995), but few implement standardized, agency-wide protocols that are applied at multiple spatial scales which would facilitate assessment and monitoring, trend analyses, or data sharing among agencies. As the largest public land management agency in the U.S., managing over one million square kilometers, the Bureau of Land Management (BLM) has attempted to overcome this shortfall by developing the National Aquatic Monitoring Framework (NAMF) (BLM 2015), a component of the BLM’s Assessment, Inventory, and Monitoring (AIM) strategy (Toevs et al. 2011).

The BLM’s AIM-NAMF seeks to improve the quality, consistency, and applicability of BLM monitoring by implementing: (1) consistent, quantitative core indicators, (2) statistically-valid sampling designs, (3) electronic data acquisition and management plans, and (4) analytical tools to increase the use of monitoring data for management decisions. Given overlap among BLM data needs and those of other federal and regulatory agencies, the AIM-NAMF adopted methods from existing, widely used, monitoring protocols, such as the EPA's NRSA (USEPA 2009), BLM/USFS’s Aquatic Riparian Effectiveness Monitoring Program (Reeves et al. 2003), BLM/USFS’s PacFish InFish Biological Opinion Effectiveness Monitoring Program (PIBO) (Kershner et al. 2004), and BLM’s Multiple Indicator Monitoring (MIM) (Burton et al. 2011) (see methods for more detail). Such collaborations will allow the BLM to leverage existing or new data collected on public lands, utilize reference data networks, and use previously developed analytical tools for making lotic condition determinations, all contributing to a more efficient and effective use of tax payer dollars.

The BLM’s mission is to “sustain the health, diversity, and productivity of America’s public lands for the use and enjoyment of present and future generations." This mission is dictated by the Federal Land Policy and Management Act of 1976 (FLPMA), which aims to ensure the sustainability of the BLM’s multiple use mandate through the inventory and monitoring of resource condition and trend, among other mandates. The rubric for assessing ecosystem health is the BLM’s Fundamental of Land Health (43 CFR 4180.1), which for lotic systems include standards for stream channel form and function, water quality, riparian areas, and biodiversity (Table 1). Lotic conditions are to be assessed for these standards to ensure the sustainability of permitted activities such as cattle and sheep grazing, recreation, and oil and gas development and to inform the adaptive management process and potential restoration needs.

Table 1. BLM stream and riparian Land Health Fundamentals and Standards for northeastern California and northwestern Nevada (43 CFR 4180.1).

Fundamentals of Land Health	Fundamental description	Land Health Standards
Stream channel form and function	Watersheds provide for the proper infiltration, retention, and release of water appropriate to soil type, vegetation, climate, and landform to provide for proper nutrient cycling, hydrologic cycling, and energy flow.	Streams (Standard 2): Stream channel form and function are characteristic for the soil type, climate, and landform. Riparian (Standard 4): Riparian areas are in proper functioning condition (i.e., vegetation is adequate to dissipate energy, stabilize stream banks, reduce incoming solar radiation, and filter sediment/nutrients).
Water quality	Water quality complies with state water quality standards or is making significant progress toward achieving the standards and BLM management objectives, such as meeting wildlife needs.	Water quality (Standard 3): Water has characteristics to support existing beneficial uses and complies with CWA and state standards.
Habitat quality for T&E and special status species	Habitats are, or are making significant progress toward being, restored or maintained for Federal threatened and endangered species, Federal proposed or candidate threatened and endangered species, and other special status species.	Biodiversity (Standard 5): Healthy, productive, and diverse populations of native and desired plant and animal species and their required habitats are maintained.

To date, the BLM has largely relied on lotic monitoring tools that do not address all Fundamentals of Land Health or are not implemented in a way that allows inference to all lotic resources in a given area. For example, Proper Functioning Condition (PFC) is commonly used to assess the condition of lotic systems in grazing allotments and uses qualitative assessments of riparian vegetation, hydrology, and channel morphology (Prichard et al. 1998). However, PFC does not address the BLM's biodiversity, habitat, and water quality standards. Additionally, PFC assessments do not result in quantitative baseline data for comparison at future dates (i.e., trend). Another example is the Multiple Indicator Monitoring (MIM) protocol (Burton et al. 2011), which is a quantitative assessment, but with a strong focus on indicators specific to grazing and implemented at sites that are selected to be 'representative' of conditions in a broader geographic region. Neither PFC nor MIM can be used to make inference to unsampled reaches with known levels of precision and confidence (Paulsen et al. 1998, Schreuder et al. 2001, McDonald 2012).

Here I present one of the first applications of the BLM's AIM-NAMF to determine the chemical, physical, and biological condition of lotic ecosystems across lands encompassed by three BLM field offices in northeast California (CA) and northwest Nevada (NV). AIM-NAMF specifies the use of statistically valid sample design and field indicators to be collected, but in terms of the analytical techniques to make condition determinations and determine causes for observed conditions, the analytical framework outlined is vague. I use a spatially balanced probabilistic design outlined in AIM-NAMF to: (1) estimate the percent of stream kilometers in two different condition classes (i.e., equivalent to and non-reference) for perennial, lotic systems at the BLM district and field office scale, (2) identify biologically relevant stressors, and (3) identify the potential sources of stressors to be targeted by management actions. For objectives two and three, I move beyond the scope of the BLM's AIM-NAMF to provide BLM managers a framework for how to determine and address the potential problems discovered from results of objective one.

METHODS

Study Area

Data collection occurred over ~10,000 square km of BLM lands managed by the Alturas (AFO), Eagle Lake (ELFO), and Surprise Field Offices (SFO) in northeast CA and northwest NV (Fig. 1). Within this area there are several quintessential BLM management priorities including grazing by cattle, sheep, and wild horses and burros. Additional management priorities in this region include sage-grouse, roads, wildfires, cultural resources, and energy development. The main management differences among field offices are wild horse and burro herds in SFO and ELFO, illegal marijuana cultivation along streams in AFO, and different landownership patterns. SFO and ELFO have large, contiguous swaths of public land in contrast to the patchy distribution in the AFO. The three field offices encompass four level III EPA ecoregions (Omernik 1987): Northern Basin and Range, Central Basin and Range, Sierra Nevada, and Eastern Cascades Slopes and Foothills (Fig. 2). Within the study area annual precipitation ranges from approximately 200 to 600 mm, elevation ranges from approximately 1,000 to

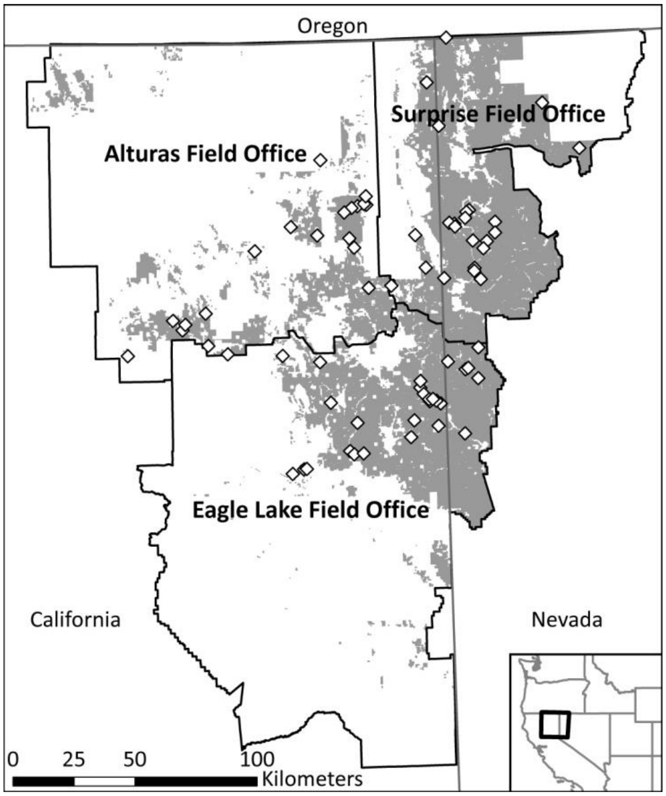


Fig. 1. BLM land ownership (grey coverage) and sample reaches (white diamonds) within the Alturas, Eagle Lake, and Surprise Field Offices.

2,400 m, and the geology is dominated by volcanic rock. The majority of BLM land (>90%) falls within the Northern Basin and Range (ELFO and SFO) and Eastern Cascades Slopes and Foothills (AFO) ecoregions. Of the four ecoregions, the Northern Basin and Range and Central Basin and Range are warmer in temperature, lower in elevation, more arid, and dominated by sagebrush steppe. In contrast, the Eastern Cascades Slopes and Foothills and Sierra Nevada are characterized by conifer forests and Sierra/Western juniper. BLM lotic resources include small, spring-fed streams, with larger streams found mostly in the Eastern Cascades Slopes and Foothills and the Sierra Nevada ecoregions.

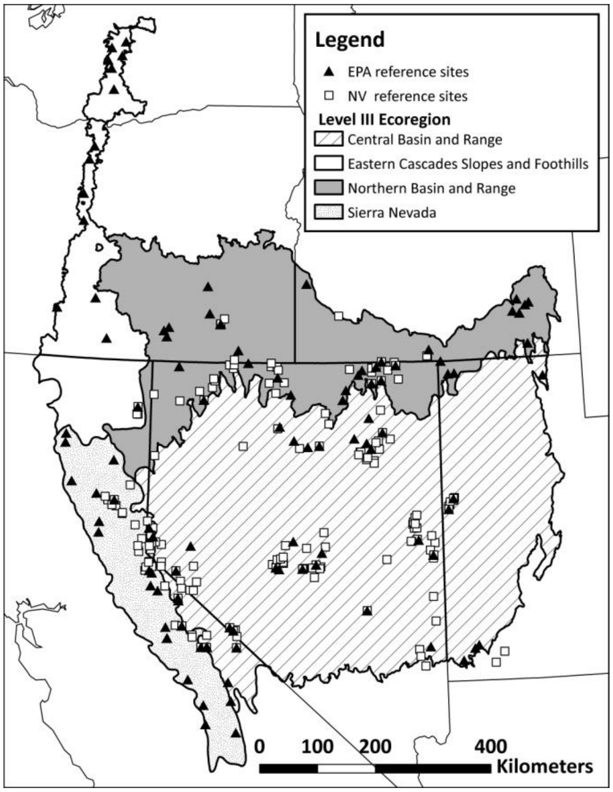


Fig. 2. Environmental Protection Agency and Nevada reference sites located within the ecoregions overlapping the study area.

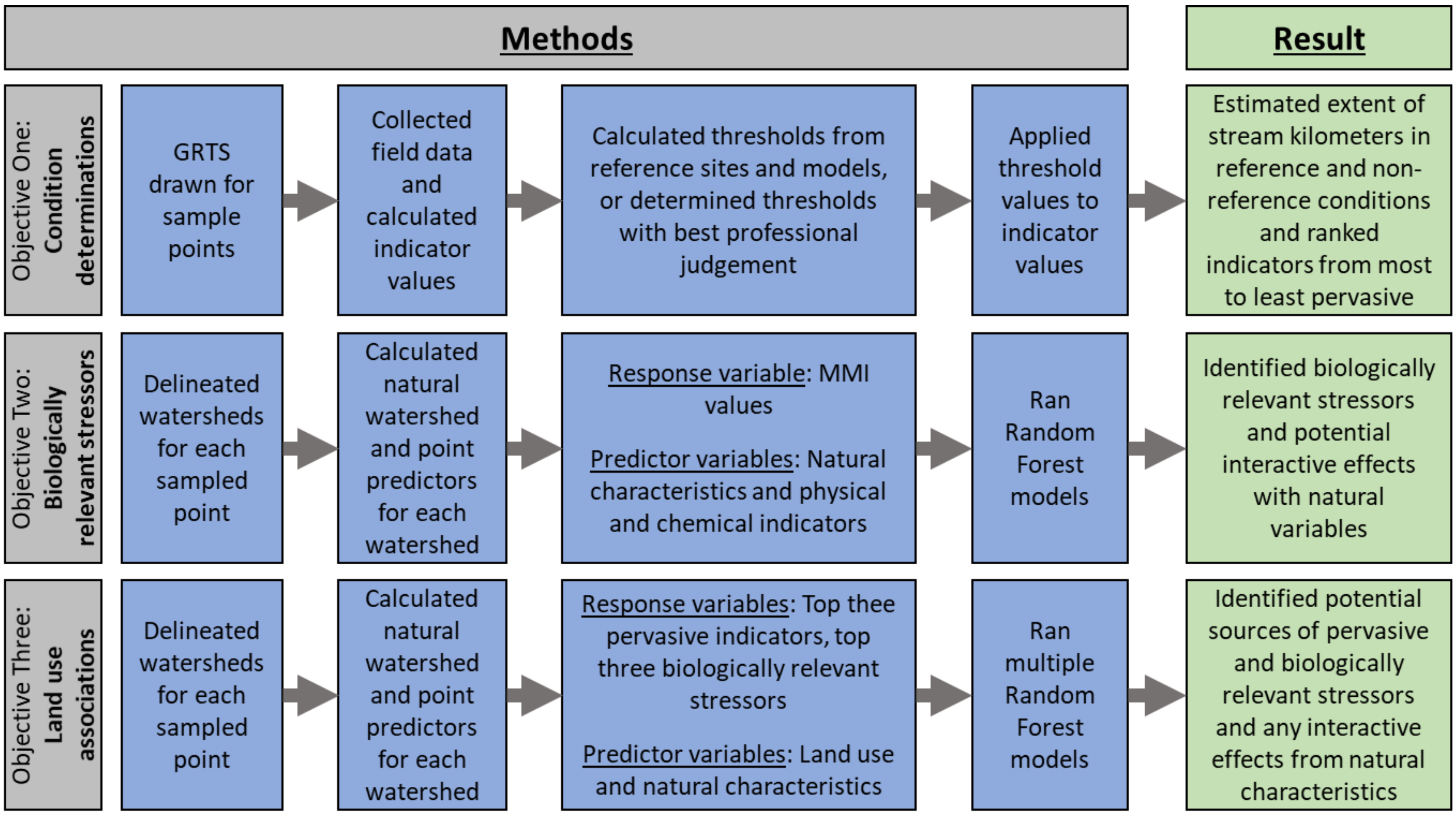
General Methods Overview

To address each of my three objectives I followed a series of field methods and geospatial and statistical analyses (Fig. 3). In the subsequent sections, I describe specific methods related to: (1) survey design, (2) field sampling, (3) deriving chemical, physical, and biological condition determinations and extent estimates, (4) obtaining watershed and site characteristics, and (5) stressor identification.

Survey Design

To select sample reaches I used a spatially balanced, stratified random sample (Generalized Random Tessellation Stratified [GRTS]; Stevens and Olsen 2004). The target population for the GRTS design was all natural, perennial streams and rivers on

Fig. 3. Methods flow chart for each of the three study objectives and the end result of the specified methods.



BLM land within the three field offices, as defined by the medium resolution (1:100,000-scale) National Hydrography Dataset Plus (NHD) (USGS 2012). Perennial flow was verified in the field during the index period (June 1^{st} – September 30^{th}) and stream reaches were sampled if they contained water throughout 50% of the sample length (USEPA 2009).

I used unequal selection probabilities, where reaches were selected in approximate proportion to the linear extent of streams within three Strahler stream order categories: small streams (1^{st} – 2^{nd}), large streams (3^{rd} – 4^{th}), and rivers (5^{th} and above). The GRTS approach is advantageous because it avoids the clustering of sites typical of simple random samples, and it can be applied to points, polygons, or polylines (Stevens and Olsen 2004, Olsen et al. 2012). Additionally, the EPA, USFS, and CA regulatory agencies commonly use GRTS designs, thus allowing data integration among monitoring efforts.

To ensure sufficient samples sizes for the scale at which land use planning decisions are made by the BLM, I stratified by field office (AFO, SFO, ELFO). I targeted 30 reaches per field office for sampling during 2013 and 2014. However, because of anticipated errors in the NHD used to define the target population, I selected oversample reaches as potential replacements. To maximize sample sizes, I also used data from an additional survey conducted within the ELFO and SFO at the grazing allotment scale (covering ~2200 km^2) that was conducted during the same time period and used the same methods described here. In total, 116 potential sample reaches, plus additional oversample reaches, were selected from the target population of 448 stream km for potential sampling (Table 2).

During field visits stream reaches were classified into one of three categories: (1) sampled, (2) inaccessible (e.g., dangerous conditions or a land owner denied access), or (3) non-target (e.g., wetland, intermittent stream, manmade canal or ditch). Once field sampling was completed, I adjusted the weights of each reach to account for the change in total stream km due to any non-target and inaccessible reaches. Initial stream reach weights (i.e., the linear extent of stream represented by any one sample) ranged from 0.7 to 9.0 stream km, but once adjusted, weights ranged from 0.4 to 7.2 stream km (Table 2).

Field sampling

Field data were collected following the AIM-NAMF protocol which is compiled from a subset of measurements from the EPA's NRSA wadeable protocol (USEPA 2009) and the BLM's MIM protocol (Burton et al. 2011). The NRSA protocol defined reach length as 40 times the average wetted width or a minimum of 150 m. Along a sample reach, 21 equally spaced transects (11 main and 10 intermediate, arranged in alternating pattern) were temporarily established perpendicular to the thalweg. Water quality parameters were collected at the center of the sample reach with a single grab sample for total nitrogen (TN) and total phosphorus (TP), and a YSI multiparameter sonde to measure pH and specific conductance. At the 11 main transects I: (1) collected benthic macroinvertebrates with a Surber sampler (0.093 m^2) fitted with a 500 µm net for total sampling area of 1.02 m^2, (2) visually estimated percent cover for instream habitat parameters (e.g., boulders and woody debris), (3) used a spherical densiometer

Table 2. Survey design summary and post sample weight adjustments by stream order category for strata and supplemental allotment surveys. Twin Peaks and Home Camp Allotment survey values were combined with Eagle Lake and Surprise Field Office values for reporting field office level results. Weights are the linear extent of stream (km) represented by one sample.

Strata	Stream Order	Target Population (stream km)	Desired Sample Size	Initial Weight	Actual Sample Size	Total Reaches assessed*	Adjusted Weight	Estimated Sample Population
Alturas	5th and above	22.0	8	2.7	3	16	1.4	4.1
	3rd - 4th	26.7	10	2.7	10	13	2.1	20.6
	1st - 2nd	38.7	12	3.2	8	27	1.4	11.5
Total Alturas Strata		87.4	30.0		21.0	56.0		36.2
Eagle Lake Field Office	5th and above	0.0	0	0.0	0	0	0.0	0.0
	3rd - 4th	34.3	8	4.3	10	12	2.9	28.6
	1st - 2nd	50.2	10	5.0	6	10	5.0	30.1
Twin Peaks Allotment	5th and above	18.7	6	3.1	2	6	3.1	6.2
	3rd - 4th	15.2	7	2.2	4	9	1.7	6.8
	1st - 2nd	16.1	7	2.3	5	10	1.6	8.0
Total Eagle Lake Strata		134.5	38.0		27.0	47.0		79.7
Surprise Field Office	5th and above	5.0	2	2.5	2	2	2.5	5.0
	3rd - 4th	63.9	10	6.4	8	13	4.9	39.3
	1st - 2nd	144.8	16	9.0	5	20	7.2	36.2
Home Camp Allotment	5th and above	0.0	0	0.0	0	0	0.0	0.0
	3rd - 4th	2.3	2	1.2	0	7	0.0	0.0
	1st - 2nd	12.8	18	0.7	7	33	0.4	2.7
Total Surprise Strata		228.8	48.0		22.0	75.0		83.2
Total all Strata		450.7	116.0		70.0	178.0		199.1

*This value represents the sum of sampled, non-target, and inaccessible stream reaches.

to estimate midstream overhead cover, and (4) visually estimated percent cover for multi-layered streamside vegetation. At all 21 transects I measured: (1) bankfull depth and the depth of the first flat depositional feature above bankfull (i.e., incision height) (See USEPA 2009 for specific sampling methods on above measurements), (2) size of 10 substrate particles from the active channel for a total of 210 particles, with a minimum of 105 particles from within the wetted width (modified USEPA 2009 method), and (3) bank stability features (i.e., slump, slough, fracture, eroding, and absent) (See Burton et al. 2011 for specific sampling methods). All field measurements were taken at baseflow conditions during the index period.

Analyses

Condition determinations – I calculated twelve indicators to represent chemical and physical stressors, and biological condition of streams (Table 3). These indicators address BLM policy (Table 1, BLM Land Health Standards [43 CFR 4180.1]), are used by state and federal regulatory agencies in implementing the CWA, and together describe the proper functioning of lotic systems. To determine the condition of the computed indicator values for each sample reach, I employed several methods including comparing the computed indicator values to: (1) values predicted by site-specific empirical modeling (e.g., Olson and Hawkins 2012, 2013, Vander Laan and Hawkins 2014), (2) thresholds set based on the distribution of values at regional reference (least disturbed) sites (e.g., Stoddard et al. 2005b, Paulsen et al. 2008), (3) national standards, and (4) thresholds set by best professional judgment (Table 4 and Table 5).

Table 3. Minimum, maximum, and median indicator and natural values among sampled reaches. Watershed area is the only non-field based measurement. Nitrogen, phosphorus, and specific conductance values are reported as observed minus expected (O-E) values (excess nutrients or specific conductance beyond predicted values).

Indicator	Minimum	Maximum	Median
O-E total Nitrogen (μg/L)	-179.5	10878.7	312.5
O-E total Phosphorus (μg/L)	-40.7	1154.8	12.6
O-E specific conductance (μS/cm)	-180.2	445.9	0.3
pH (SU)	4.9	9.7	8.3
Bank stability (%)	10	100	100
Floodplain connectivity	-1.0	0.3	-0.5
Riparian canopy cover (%)	0.0	94.1	23.1
Excess fine sediment (%)	3.3	99.0	42.9
Instream habitat complexity	0.0	1.0	0.3
NV MMI score	16.4	65.5	43.4
Invasive benthic macroinvertebrate	Presence	Absence	NA
Riparian habitat complexity	0.3	1.8	0.9
Bankfull Width (m)	0.6	27.4	3.8
Water surface slope (%)	0.6	14.8	2.1
Watershed Area (km^2)	0.6	9929.1	72.8

To estimate biological condition I considered three different benthic macroinvertebrate indices because the study area is geographically split between CA and NV: NV multimetric (MMI) and observed to expected (O/E) indices (Vander Laan and Hawkins 2014), and an MMI and O/E hybrid index created for the state of CA (Mazor et al. 2013). MMI and O/E indices are the two most common methods used to quantify the biological condition of streams and rivers with benthic macroinvertebrates (Bonada et al. 2006, Cao and Hawkins 2011). However, MMI and O/E indices each measure different aspects of biological condition. An MMI measures overall biological integrity, whereas O/E measures taxonomic completeness. Specifically, an MMI is a compilation of metrics representing compositional, functional, and tolerance characteristics of macroinvertebrate

Table 4. Water quality and biological indicator thresholds and methods used to assign condition classes of reference, non-reference, and undetermined. The undetermined category was only applied to biological condition estimates and the condition of chemical and physical stressors was only classified as non-reference.

Indicator	Condition determination method	Thresholds[1]
Macroinvertebrate biological integrity	Equivalence test	>=47.0 <=44.6
Invasive benthic macroinvertebrates	Presence / absence	NA
Total nitrogen (µg/L)	95th percentile of model error[2]	114.7 µg/L > model prediction
Total phosphorous (µg/L)	95th percentile of model error[2]	21.3 µg/L > model prediction
Specific conductance (µS/cm)	95th percentile of model error[3]	53.7 µS/cm > model prediction
pH (SU)	National standards	<6.5 and >9.0
Banks Stability (%)	Best professional judgement[4]	<60%

[1]Thresholds listed from least to most disturbed conditions
[2]Olson and Hawkins 2013
[3]Olson and Hawkins 2012
[4]Developed in consultation with BLM and USFS resource specialists

assemblages (Stoddard et al. 2008). O/E indices measure taxonomic completeness, computed as the ratio of the observed macroinvertebrate taxa to the taxa expected to occur at a site in the absence of anthropogenic impacts (Hawkins 2006). All considered indices use site-specific empirical models, built from a network of reference sites, to make predictions of the biota or metrics expected to occur at a site in the absence of anthropogenic impacts. Thus the indices account for natural variation in the biological potential of each stream reach.

To determine which network of reference sites, NV or CA, was most representative of the environmental variability of my test sites (i.e., sampled reaches), I used Nonmetric Multidimensional Scaling (NMDS) ordinations. Specifically, I ordinated

Table 5. Instream and riparian reference threshold values for the three EPA level III ecoregions used to assign the condition class of non-reference conditions for indicators for which the regional reference approach was applied. Values for only three of the four ecoregions overlapping the study area are presented because all sample sites in the Central Basin and Range ecoregion were intermittent (i.e., non-target) at the time of site visits.

		EPA hybrid ecoregion		
Indicator	Percentile of reference distribution	Eastern Cascades Slopes and Foothills	Sierra Nevada	Northern Basin and Range
Fine sediment (% < 2 mm)	90th	>39.0	>36.0	>69.0
Instream habitat complexity	10th	<0.36	<0.22	<0.11
Floodplain connectivity	90th	>0.12	>-0.52	>0.22
Riparian complexity	10th	<0.78	<0.88	<0.68
Riparian canopy cover	10th	<29.71	<22.89	<6.87

reference and test sites in environmental space, defined by the predictor variables used in each respective model, independently for the NV and CA reference networks (Table A-1). The resulting ordinations were examined to determine the degree of overlap between reference and test sites for their naturally occurring features. Test sites falling outside of the environmental space defined by a reference network were considered outliers. The reference network with the least number of associated outliers was considered most representative of the environmental variability of my test sites and as such, the model built with this reference network would be most appropriate to use for my test site locations. I also removed three test sites from the biological integrity analysis (below) because I identified these test sites as being outside of the experience of the models.

Due to the results of the above analyses (See Results), I only evaluated the NV MMI and NV O/E index to determine which one was more appropriate based on the following criteria: (1) index performance (i.e., accuracy, precision, bias, sensitivity, and responsiveness) (Vander Laan and Hawkins 2014), (2) BLM policy and objectives, and (3) benthic macroinvertebrate community assembly rules in the context of the physiographic conditions of the study area (Sada et al. 2005, Rader et al. 2012, Vander Laan and Hawkins 2014). Ultimately, I used the NV MMI to quantify the biological condition of stream reaches. Specifically, I used standardized operational taxonomic units (Hawkins et al. 2000), defined for the NV MMI, and a fixed count of 300 individuals (Ostermiller and Hawkins 2004) to calculate seven MMI metrics: insect richness, Ephemeroptera relative abundance, Shannon diversity, collector-filterer relative abundance, Plecoptera relative abundance, non-insect richness, and clinger richness. Each metric value was rescaled to a value between 0 and 100 and a final MMI score was calculated by taking the average of all seven metrics (Vander Laan and Hawkins 2014).

Sampled reaches were classified with 95% confidence as equivalent to the reference distribution (i.e., reference condition) or dissimilar to the reference distribution (i.e., degraded condition) (Vander Laan and Hawkins 2014). MMI values between the two thresholds (i.e., degraded and reference condition) could not be classified into either of these two categories with confidence and were considered undetermined. Although the MMI assesses biological integrity, it does not explicitly state the presence of invasive macroinvertebrate species. As such, independent of the MMI I assessed invasive benthic macroinvertebrates on a presence or absence basis to address this aspect of biological condition. Specifically the invasive benthic macroinvertebrates I considered were: (1) any individual in the crayfish family Cambaridae, (2) Asian clam (*Corbicula fluminea*), (3) zebra mussels (*Dreissena polymorpha*), (4) quagga mussels (*Dreissena rostriformis*), (5) New Zealand mud snail (*Potamopyrgus antipodarum*), and (6) red-rimmed melania snail (*Melanoides tuberculatus*).

Site specific predictive Random Forest models for nutrients (Olson and Hawkins 2013) and specific conductance (Olson and Hawkins 2012) were used to determine the condition of chemical parameters. These models used natural characteristics of the sample point and its watersheds to predict the naturally occurring value of nutrients and specific conductance at each sample point in the absence of anthropogenic impacts. I then subtracted the observed field values from modeled values for each site. The differences falling outside of the 95th percentile of model error were classified as non-reference (i.e., degraded) (Table 4). These empirical models were the preferred method for establishing condition thresholds because they account for natural environmental gradients among sites (Hawkins et al. 2010, Hill et al. 2013), however several stressors lacked such models.

For stressors lacking empirical models, I used the range of variability among regional reference sites (Fig. 2) to set thresholds for classifying non-reference conditions (e.g., Stoddard et al. 2005b, Paulsen et al. 2008). The regional reference approach relies on networks of sampled reference sites located within a relatively homogenous physiographic region (e.g., Omernik level III ecoregions) to establish the expected range of conditions in the absence of anthropogenic impacts (Stoddard et al. 2006, Hawkins et al. 2010). In contrast to the empirical model approach, the regional reference approach accounts for variation among ecoregions, but not among sites within an ecoregion. For stressors lacking empirical models in this study, I used the EPA's NRSA and Environmental Monitoring and Assessment Program Western streams and rivers (EMAP-West) reference sites located in level III ecoregions encompassed by the three field offices to establish the expected range of conditions in the absence of anthropogenic impacts (Omernik 1987). Sample sizes of reference sites ranged from 13 to 32 sites per stressor and ecoregion (Fig. 2). Thresholds were established at the extremes of reference distributions (i.e., 10% or 90%) to identify non-reference conditions for floodplain connectivity, instream complexity, canopy cover, riparian complexity, and percent fine sediment (Paulsen et al. 2008; Table 5). The choice to use the 10% and 90% of the reference distribution was a management decisions made by the BLM resource staff. Bank stability and pH lacked both empirical models and regional reference values, so I collectively set threshold values to establish non-reference conditions with BLM staff based on best professional judgement (based on experienced resource specialists of the BLM and Forest Service) and the EPA's national recommended aquatic life criteria table setting national standards respectively (Table 5).

Condition Extent Estimates – I computed the extent of stream kilometers in reference and non-reference condition for MMI scores and the extent in non-reference for chemical and physical stressors. Extent estimates were calculated for the entire study area and for each strata in the study design by summing the adjusted weights of each sampled reach in a condition class and dividing by the sum of sampled weights in all condition classes for each scale (i.e., entire study area, AFO, ELFO, and SFO). By summing only the adjusted weights for the sampled reaches, I made inference to the population of sampled streams, but not the population of inaccessible or non-target streams. To calculate extent estimates and 90% confidence intervals, I used the EPA's 'spsurvey' package version 2.6 with the Horvitz-Thompson variance estimator (Kincaid et al. 2013) in R statistical software version 2.15.0 (R Core Team 2012). I chose 90% confidence intervals, with consultation with BLM resource staff, to balance type I and type II errors, while recognizing that the consequence of a false positive in the context of natural resource management for the BLM has relatively low consequences at this scale of analysis compared to false negatives.

Obtaining watershed and site characteristics – Natural (e.g., climate, soil type, geology) and anthropogenic (e.g., road density, grazing intensity, land cover) characteristics (Table A-1) of sampled reaches and their respective watersheds were used: (1) to determine the appropriate biological index, (2) as predictor variables for site-specific empirical models, (3) to model spatial variability in biological conditions as a function of stressors and physiographic conditions (i.e., biologically relevant stressors), and (4) to model spatial variability in both biologically relevant and highly extensive stressors as a function of land use and physiographic conditions (Table A-1). Watershed boundaries were delineated upstream of sampled reaches with the multi-watershed delineation (MWD) tool (Chinnayakanahalli et al. 2006) in ArcMap 9.3. I then used ArcMap 10.1 to quantify natural and anthropogenic characteristics at either the point or watershed scales (Carlisle et al. 2009; See Appendix).

Stressor identification – I moved beyond the scope of the BLM's AIM-NAMF and developed Random Forest (RF) models to quantify two types of relationships: (1) spatial variation in biological condition as a function of measured chemical and physical stressors and natural characteristics, and (2) spatial variation in both the top three biologically relevant stressors (identified in the first RF model) and the three chemical and physical stressors with the greatest extent of streams km in non-reference condition as a function of anthropogenic land use. The objective of the latter analysis was to identify likely sources of stressors (See model details below).

RF models fit many regression or classification trees with bootstrapped samples of the data and a random subset of predictors at each split in the tree. The results from all trees are then averaged to make predictions (Breiman 2001, Liaw and Wiener 2002, Cutler et al. 2007). RF models are increasingly used for modeling complex biological responses (e.g., Carlisle et al. 2009, Chinnayakanahalli et al. 2011, Vander Laan et al. 2013) as they have been shown to outperform other parametric and non-parametric techniques, can be used with both categorical and continuous data, and are resistant to overfitting (Prasad et al. 2006, Peters et al. 2007, Olson and Hawkins 2012). I ran RF in regression mode in R statistical software version 2.15.0 (R Core Team 2012) with the randomForest 4.6-7 package using the default number of trees (500). Model performance was assessed using the percent variance explained, which is an internal cross-validated metric defined as 1 - (mean squared error)/(variance (response)), and is analogous to an r-squared (Pang et al. 2006). I assessed variable importance by evaluating the percent increase in mean square error (MSE) following removal of each predictor variable, with higher MSE values indicating a greater decrease in model accuracy (Pang et al. 2006).

To identify field collected indicators associated with degraded biological conditions (hereafter referred to as biologically relevant stressors), I built RF models with NV MMI scores as my response variable and both natural watershed characteristics and field collected indicators as predictor variables. Natural characteristics were included to identify any unknown bias in the NV MMI model or potential interacting effects between biologically relevant stressors and natural variables. Model development was an iterative process where all stressors and a subset of natural watershed characteristics were included as predictor variables. I then iteratively removed the least important predictor variables until the percent variance explained was maximized. I assessed model performance by comparing the % variance in MMI scores explained by the RF models to the maximum possible variance explained given the variability in MMI reference scores due to index error. The maximum possible variance explained was calculated as:

$$100 \times \frac{S{:}N}{(S{:}N) + 1}$$

where *S:N* was the signal-to-noise ratio. The signal was the variance among all MMI scores (i.e., reference and sampled test sites) and the noise was the variance among reference sites used to develop the MMI (Vander Laan et al. 2013).

To identify possible stressor origins, I developed RF models relating among-site variation in stressor values to anthropogenic land uses. Specifically, I modeled the top three biologically relevant stressors and the three stressors with the greatest extent of stream km in non-reference condition as a function of anthropogenic land uses (e.g., road density, agriculture) and natural watershed characteristics. I used the same iterative methods as above to identify the best model and determine the most important predictor variables to each response variable. To assess RF model performance I calculated the maximum possible variance explained using same methods as above for the RF models where water quality was the response variable. I then compared the % variance in water quality scores explained by the RF models to the maximum possible variance explained. For those stressors that lacked empirical models the maximum possible percent variance was 100%.

RESULTS

Target population

A total of 70 reaches were sampled out of 178 base and oversample reaches visited across the three field offices (Table 2). Sampled reaches were used to compute condition estimates for 199 stream km, which represented 44% of the 448 km initially identified by the NHD. I was unable to make condition estimates for 71 km (16%) due to inaccessibility and thus the condition of these stream segments was considered unknown. Non-target reaches accounted for 178 km (40%), largely due to intermittent flow, and were excluded from condition extent estimates. Of the 199 perennial stream km, AFO contained 36 km, ELFO 80 km, and SFO 83 km.

<u>Macroinvertebrate condition extent estimates</u>

The NV MMI was used to derive macroinvertebrate biological condition estimates because: (1) the NV reference network was more representative of the environmental variability of sampled reaches compared to the CA reference network (Fig. 4), (2) the NV MMI outperformed the NV O/E index (Table 6), (3) an MMI directly addressed BLM policy and objectives, and (4) an MMI was better suited for the study area given the occurrence of highly isolated stream networks which can confound the species distribution models underlying O/E indices (Vander Laan and Hawkins 2014).

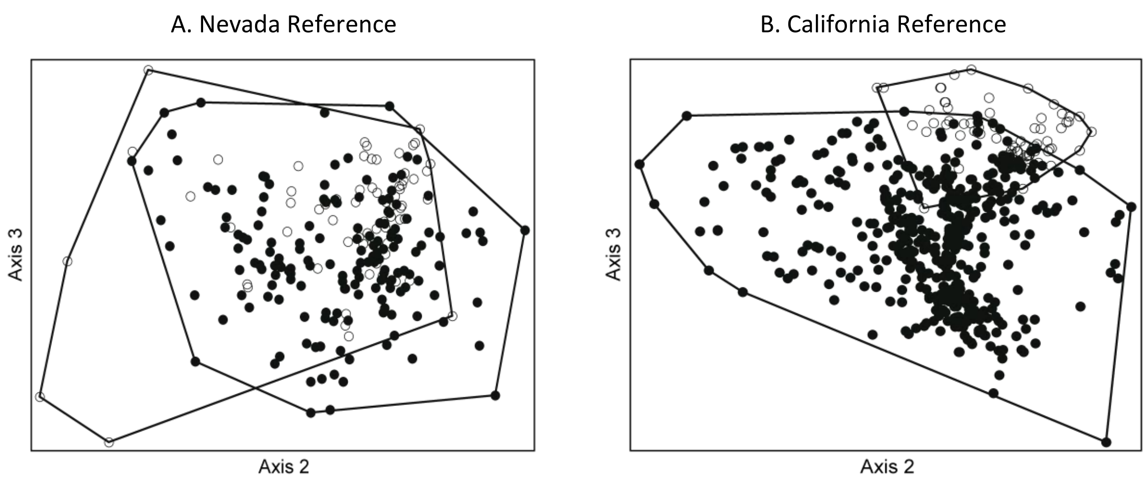


Figure 4. NMDS ordinations of sampled reaches (open circles) and reference sites (closed circles) for the Nevada (A) and California (B) bioassessment models. Environmental space is defined by the predictor variables used in each respective model, independently for the Nevada and California reference networks. The polygons within each ordination are the convex hull areas or extent of environmental heterogeneity for the sampled and references reaches.

Across all three field offices NV MMI scores indicated that 90 km (45%) were in non-reference condition, 77 km (39%) were in reference condition, and 18 km (9%) were

Table 6. Comparison of Nevada multimetric index (MMI) and observed divided by expected (O/E) index performance measures. RF % Var = percent variance of reference site scores associated with natural gradients.

Index	Reference samples (n=165)			Test samples (n=374)			
	Mean index score	Precision CV[2]	Bias RF % Var[3]	Mean index score	Responsiveness[4] Mean difference	t-Value	Sensitivity % degraded[5]
O/E-0	1	0.21	0.1	0.93	0.07	3.4	17.6
O/E-5	1.06	0.25	0	0.97	0.09	3.8	15.2
MMI[1]	1	0.11	0	0.93	0.07	5.4	29.9

Table modified from Vander Laan and Hawkins 2014
[1]MMI scores standardized by dividing by reference score mean
[2]Precision is the coefficient of variation among reference scores.
[3]Bias is the percent of variation in reference index scores explained by natural environmental predictors used to build the model.
[4]Responsiveness is the different between reference and test site (predetermined non-reference sites) index scores.
[5]Sensitivity is the percent of predetermined non-reference sites, used as test sites, correctly determine by the index to be in non-reference condition, higher values are better.

unable to be classified with confidence (undetermined) (Fig. 5). Among individual field offices, the AFO and ELFO had the lowest percentage of stream km in non-reference condition 38% and 38%, respectively. In contrast, the SFO had the highest percentage in non-reference condition (56%).

Stressor extent estimates

The stressors with the greatest extent of stream length in non-reference condition for all three field offices combined were TN (68%), riparian cover (43%), and TP (37%), whereas the least pervasive stressors were bank stability (10%), floodplain connectivity (4%), and benthic invasives (3%) (Fig. 6). However, the stressors with the greatest relative extent varied among strata. For AFO, instream complexity was the most extensive stressor, with 63% of stream length in non-reference conditions, followed by 54% for TN, and 44% for riparian canopy cover. In ELFO, 60% of stream length was in non-reference for TN, 39% for riparian canopy cover, and 35% for pH. In SFO, 82% of

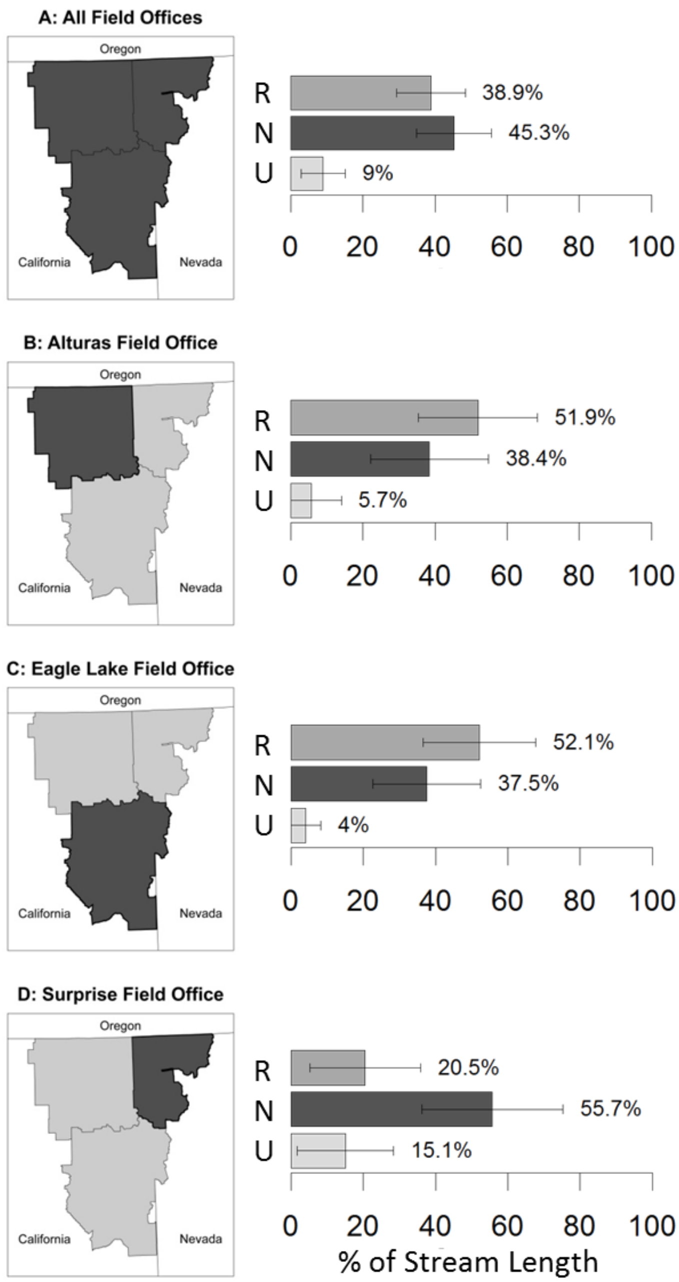


Fig. 5. Biological condition extent estimates (± 90% confidence interval) based on NV MMI scores for all field offices combined (A) and the Alturas (B), Eagle Lake (C), and Surprise (D) Field Offices. Medium grey bars represent sites equivalent to reference condition (R), dark grey represents non-reference (N), and light grey represents undetermined conditions (U). Percentages do not add to 100 because three sites were omitted that did not fit the experience of the NV model.

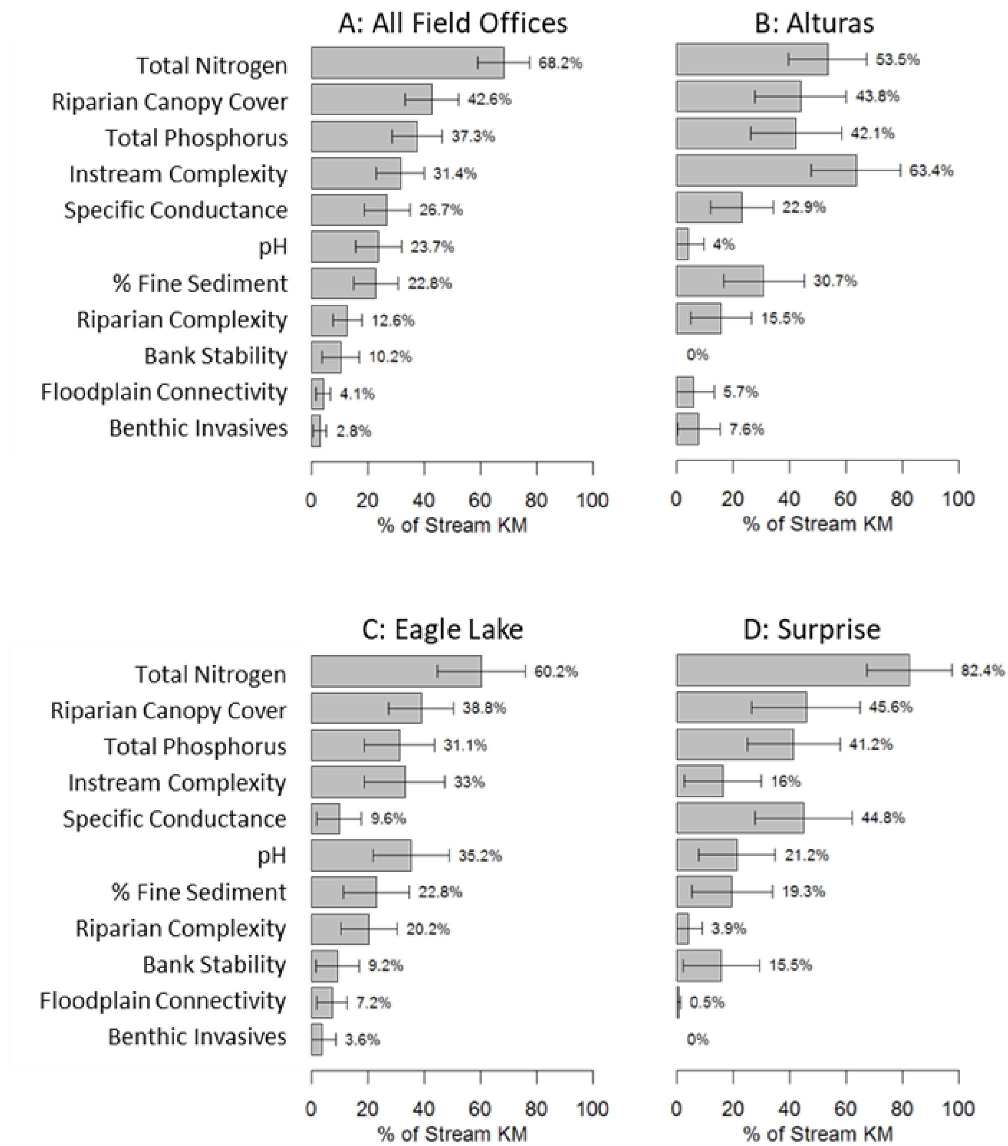


Fig. 6. Relative extent estimates (±90% confidence intervals) of chemical and physical stressors in non-reference conditions for all three field offices combined (A) and individually for Alturas (B), Eagle Lake (C), and Surprise (D) Field Offices.

stream length was in non-reference condition for TN, 46% for riparian cover, and 45% for specific conductance.

Biologically relevant stressor identification

Three stressors and two natural watershed characteristics accounted for 28% of the variance in MMI scores out of a total possible variance of 71%: watershed area, riparian complexity, TN, intermittent stream density, and TP (in order of importance; Fig. 7). Threshold responses characterized the relationship of MMI scores with all predictors. MMI scores drastically decreased when TN concentrations were 325 µg/L above predicted natural conditions and TP concentrations exceeded 80 µg/L of predicted natural conditions (Fig. 8). In contrast, MMI scores abruptly increased when riparian complexity values exceeded 0.9 (unitless) and watershed areas were greater than > 10 km^2. Several predictors exhibited interactive effects, for example the lowest MMI scores were observed among small watersheds with high nutrient concentrations or reaches with reduced riparian complexity and high nutrient concentrations (Fig. 9).

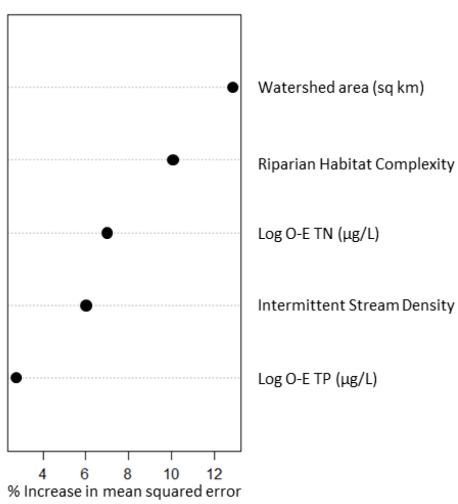


Fig. 7. Variable importance plot from Random Forest model identifying biologically relevant stressors.

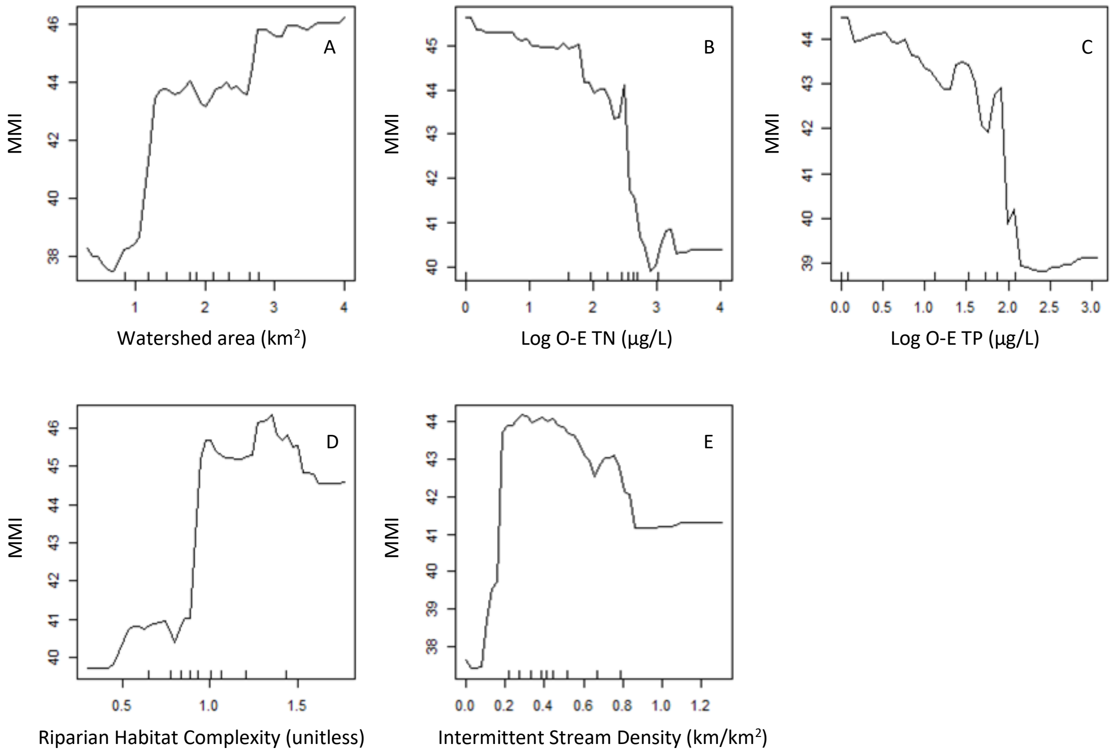


Fig. 8. Partial dependence plots of sample reach MMI scores as a function of: watershed area (A), log of O-E TN (B), log of O-E TP (C), riparian habitat complexity (D), and intermittent stream density (E). These predictor variables were identified as biologically relevant from Random Forest models and are displayed in order of importance. Plots show MMI scores for each predictor variable after averaging out the effect of all other predictors in the top models. Rug plots (vertical lines extending upward from the x-axis) indicate deciles of data for each predictor variable.

Sources of stressors

Land use and natural watershed characteristics accounted for 11-26% of the variance in biologically relevant and spatially extensive stressors (Table 7). Nutrient exceedances were most strongly associated with grazing and natural watershed characteristics, whereas riparian alteration was not strongly associated with land uses. For example, excess TN and TP were both positively related to the amount of long term grazing within watersheds. A positive, threshold response occurred for TN and TP when

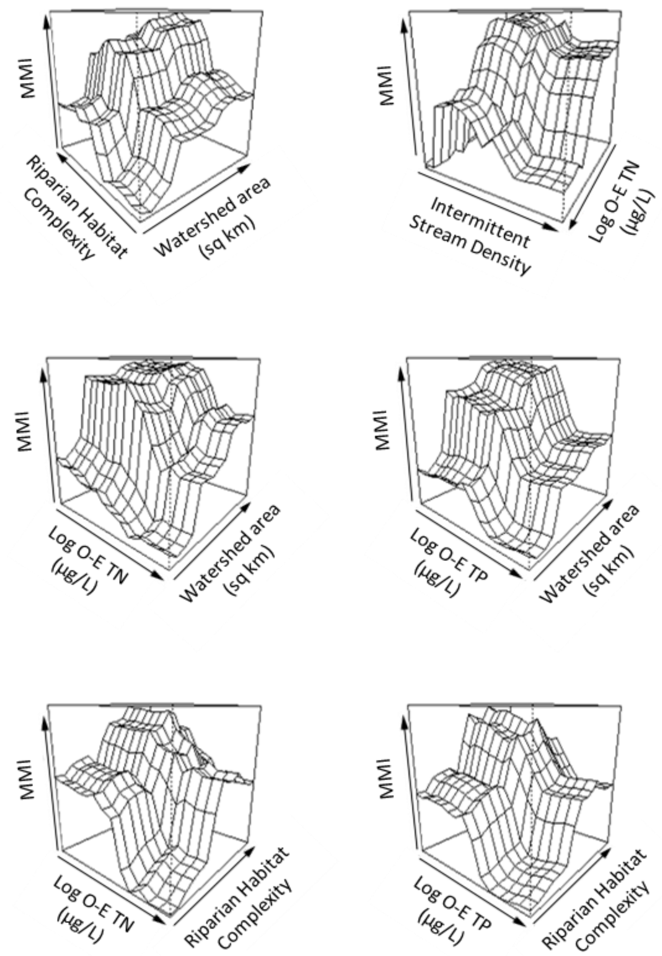


Fig. 9. Bivariate partial dependence plots showing the joint effects of stressors and natural variables on MMI scores.

grazing exceeded 0.5 times the permitted utilization for both short and long term grazing (Fig. 10). In contrast, relationships of TN and TP with anthropogenic variables such as road density and stream crossings and natural variables such as slope and drainage density were opposite of what one would expect and largely uninterpretable.

Among site variation in both riparian cover and riparian complexity was associated with only natural watershed characteristics. Alterations in riparian complexity

Table 7. Top Random Forest models for the relationships between anthropogenic land uses and both biologically relevant (O-E TN, O-E TP, and Riparian Complexity) and spatially extensive (O-E TN, O-E TP, and Riparian Cover) stressors. % var is the % variation in each stressor accounted for by the land use and natural predictors listed. Predictors (See Appendix) listed from most important to least important in terms of % variance explained.

Stressors	% var	Predictors
O-E TN	11 out of 94.7	Grazing 3 years prior to sampling (+), watershed slope (U-shaped), grazing one year prior to sampling (+)
O-E TP	24 out of 92.0	Stream density (+), grazing 3 years prior to sampling (+), density of road-stream crossings (+)
Riparian Complexity	26	Average monthly discharge (+), strahler steam order (bell-shaped)
Riparian Cover	23	Intermittent stream density (-), minimum average air temperature (+)

was the most predictable indicator and was associated with average monthly discharge and stream order. The variation in riparian cover was associated with intermittent stream density and the average minimum air temperature of the watershed.

DISCUSSION

The BLM manages more public lands than any other agency in the U.S. and is required to manage the National System of Public Lands under a multiple-use mandate (43 U.S.C. §1701 et seq.). The successful implementation of this mandate requires timely and accurate information regarding resource condition and trends to ensure that permitted uses (e.g., livestock grazing, recreation, mineral extraction) are managed in such a way that the health, diversity, and productivity of public lands are sustained for present and

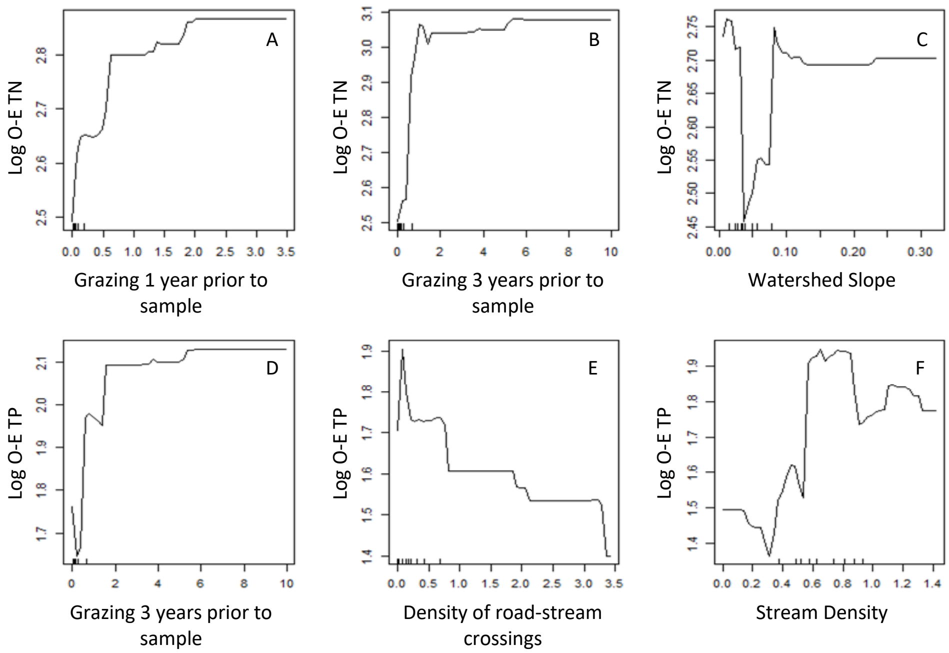


Fig. 10. Partial dependence plots of sample reach O-E TN (A-C) and O-E TP (D-F) scores (logged) as a function of: grazing 1 year prior to sampling (A), grazing 3 years prior to sampling (B), watershed slope (C), Grazing 3 years prior to sampling (D), density of road-stream crossings in the watershed (E), and stream density of the watershed (F). These predictor variables were identified as the likely sources of excess nutrients from Random Forest models. Plots show log of excess nutrient values for each predictor variable after averaging out the effect of all other predictors in the top models. Rug plots (vertical lines extending upward from the x-axis) indicate deciles of data for each predictor variable.

future generations. Application of the NAMF field methods provided the necessary field based monitoring data for BLM management decisions (but see data and capacity limitations section below), but until now there was no proposed analytical framework for how to use and interpret the resulting monitoring information to inform management. Below I discuss the observed results and highlight both the application and research challenges that can limit effective implementation of resource monitoring and adaptive management strategies necessary to meet federal regulations.

Condition Estimates

The types of stream degradation I observed are comparable to those identified in California's statewide Perennial Streams Assessment (PSA), while the magnitude of such impacts was lower. Ode et al. (2009) found approximately 70% of stream kilometers in the Desert-Modoc region (the PSA reporting area overlapping the present study area) in degraded biological condition, compared to 45% of stream km in my study. Similarly, TN and TP were among the most ubiquitous stressors affecting 70% and 80% of stream kilometers, respectively (Ode et al. 2009), compared to 68% and 37% of stream km in my study respectively.

Differences in the extent of stream kilometers experiencing degraded biological condition and excessive nutrient loading between studies likely results from disparate target populations. The PSA encompassed all lands and streams regardless of ownership, whereas my study focused exclusively on BLM lands and defined the target population as all stream and rivers on BLM land. The type, distribution, and intensity of anthropogenic impacts is not uniform across the Desert-Modoc region, especially for public lands. For example, the watershed area upstream of my sampled reaches did not contain more than 9% agriculture, whereas 35% of the reaches sampled in the PSA had >50% agriculture in the upstream watershed. Agriculture changes runoff patterns, increases nutrient loading, and alters physical attributes of streams, all of which have the potential to decrease biological integrity (Allan 2004, Matthaei et al. 2010).

Biologically relevant stressors

Excessive nitrogen and phosphorous and reduced riparian complexity were the most biologically relevant stressors, a result that is consistent with the findings of other studies linking macroinvertebrate condition to instream stressors (Miltner and Rankin 1998, Townsend et al. 2008, Van Sickle and Paulsen 2008). TN and TP are necessary nutrients for macroinvertebrate growth and survival, but increased concentrations can have negative effects on macroinvertebrate assemblages (Perrin and Richardson 1997, Bourassa and Cattaneo 1998, Miltner and Rankin 1998, Yuan 2010). For example, excess TN or TP can increase periphyton biomass causing changes to dissolved oxygen concentrations, pH levels, and other factors that can alter macroinvertebrate richness and composition (Hart and Robinson 1990, Delong and Brusven 1998, Dodds and Welch 2000, Wang et al. 2007, Hayashi et al. 2012). Increased nutrients also accelerates leaf litter breakdown rates, which alters food quality and the timing of food availability, which in turn can alter growth rates and food web dynamics (Robinson and Gessner 2000, Niyogi et al. 2003, Cross et al. 2005, Greenwood et al. 2007). Paulsen et al. (2008) concluded that reducing nutrient loading to streams could have the single largest positive effect on macroinvertebrate biological condition throughout the nation.

Similarly, there are numerous direct and indirect mechanisms by which degraded riparian areas can negatively affect macroinvertebrates (Delong and Brusven 1994, Naiman and Décamps 1997, Rios and Bailey 2006). I found that biological integrity was most degraded when degraded riparian complexity and excess nutrients co-occurred. Riparian areas act to filter nutrients, so when riparian vegetation is degraded more nutrient loading can occur. Cattle, sheep, horse, and burro grazing were ubiquitous throughout the study area and were the likely source of degraded riparian vegetation and excess nutrients (see later discussion section). Additionally, reduced riparian vegetative cover can increase incoming solar radiation, which can also interact with excess nutrients to exacerbate stream algal blooms. It is common to find multiple stressors simultaneously influencing freshwater systems and it is important to understand their interactions or additive effects on biological conditions (Folt et al. 1999, Culp et al. 2000, Townsend et al. 2008). Alternatively, because riparian complexity was more strongly associated with natural environmental factors than anthropogenic impacts, the relationship with MMI scores might reflect macroinvertebrate responses to natural environmental gradients of riparian complexity. Although the reference sites used to develop the NV MMI model were more representative than those for the California model for my test sites, some discrepancies did exists (see discussion below).

In general, MMI scores exhibited threshold responses to nutrient loading and riparian degradation. The threshold concentration of TN (~500 μg/L) and TP (120 μg/L) associated with biological degradation are consistent with other studies, which found nutrients to alter or degrade macroinvertebrate assemblages with thresholds ranging from 590 to 2830 μg/L for TN and 21 to 150 μg/L for TP (Wang et al. 2007, Weigel and Robertson 2007, Chambers et al. 2012). In contrast to the biologically relevant thresholds I identified for macroinvertebrates, the thresholds derived from predictive models for TN and TP, and used to determine excess TN and TP for stressor extent estimates, were more than 60% lower. This discrepancy represents a management challenge for the development of thresholds for making condition determinations. Specifically, whether thresholds such as that set for nutrients should be based on what is physically achievable at a site based on local physiographic conditions versus exceedances that have deleterious biological impacts (Ice and Binkley 2003, Dodds 2007, Olson and Hawkins 2013). I chose to focus on what is achievable at a site based on predicted natural conditions because streams and rivers provide many beneficial uses and the BLM is interested in minimizing the potential for downstream or cumulative impacts such as nutrient loading. Additional, we only considered impacts to benthic macroinvertebrates, while the impact to other trophic levels, biotic assemblages, or beneficial uses might occur at lower nutrient concentrations than the threshold levels I observed for macroinvertebrates (Passy et al. 2004, Hughes et al. 2009, Marzin et al. 2012).

In contrast to the potentially over protective thresholds (from a macroinvertebrate standpoint) for excess nutrients resulting from site specific predictive models, the riparian complexity thresholds set with regional reference are potentially under protective (from a macroinvertebrate standpoint). The MMI threshold response observed for riparian complexity values was approximately equal to the largest of the three ecoregional thresholds derived from regional reference sites. This result indicates that for the two ecoregions with lower thresholds, the regional reference thresholds are potentially under protective of riparian areas from a macroinvertebrate biological condition perspective. The difference between over protective nutrient thresholds and under protective riparian thresholds likely resulted from the use of site-specific empirical models versus regional reference conditions, respectively. Specifically, if the natural variation in an indicator is not accounted for, as was the case for the regional reference condition approach for riparian complexity, the lack of precision and accuracy in setting thresholds can result in either unattainable or under protective thresholds (Olson and Hawkins 2013).

In addition to measured stressors, macroinvertebrate MMI scores were also related to natural predictors. For example, watershed area and intermittent stream density were among the top predictors of MMI scores. There are several possible reasons for these associations: (1) the NV MMI predictive model did not completely account for the full range of environmental gradients throughout the study area, (2) some context dependency is occurring, or (3) a combination of these. Vander Laan and Hawkins (2014) reported minimal bias when checking for residual variance in MMI scores as a function of natural environmental gradients (Table 6). However, discrepancies exist in the extent and type of environmental gradients used to calibrate the model and those encompassed by this study. For example, NMDS ordination showed that some sampled reaches fell outside the environmental space represented by model calibration reference sites. After examining boxplots and scatterplots of NV model predictor variables, I observed that small watersheds at low elevation were underrepresented in the model's reference sites (See Appendix B, Figs. B-1 and B-2). Therefore, biological condition estimates for low elevation, small watersheds may not be accurate. I removed the three most obvious outliers from the MMI results, however there may be remaining low elevation small watershed sites that are underrepresented, but may not be as obvious.

Alternatively, natural variables may have interactive effects with stressors. For example, small watersheds with even low levels of riparian degradation or nutrient loading had much lower MMI scores than larger watersheds with the same stressor levels. Many of the small watersheds in the study area are spring-fed headwater systems Macroinvertebrate assemblages in small spring-fed systems may have narrower environmental optima resulting from relative high degree of environmental stability (Barquín and Death 2004). This could explain my observation of even small alterations of riparian vegetation and nutrient loading having more adverse effects on stream biota for small spring-fed streams, than the same level of alterations in the larger less spring influenced watersheds. The health of small headwater streams is extremely important as they feed water, sediment, and nutrients to the rest of the dendritic system and a breach of health in a headwater system can cause subsequent degradation to higher order streams (Sidle et al. 2000, Gomi et al. 2001, Nadeau and Rains 2007).

Sources of stressors

Excess nutrients were most strongly related to the intensity and duration of cattle grazing, the predominant land use in the study area. This result suggests that to meet nutrient thresholds in the future managers in this study area should reduce or change grazing regimes. Livestock grazing is permitted on approximately 90% of public lands in the western U.S. (Kauffman et al. 1997), so it is not surprising that my results identified livestock grazing is the likely cause of region-wide excess TN and TP. Several other studies have shown that grazing increases nutrient loading, decreases riparian vegetation and shading, increases fine sediment (bed and suspended sediments), erodes banks, and has other environmental effects (Fleischner 1994, del Rosario et al. 2002, Asner et al.

2004, Beschta et al. 2012), all of which have been linked to changes in macroinvertebrate assemblages (Herbst et al. 2012).

Natural watershed characteristics alone and not anthropogenic impacts were associated with spatial variability in riparian complexity (associated with stream order and average discharge) and canopy cover (associated with intermittent stream density and minimum air temperature). The amount and timing of water availability, as well as climatic conditions can cause natural variation in riparian vegetation composition and structure (Naiman and Décamps 1997). Such results likely reflect the fact that site-specific empirical models were not available for riparian complexity or canopy cover and thus natural environmental gradients overwhelmed any potential signal from land uses. Although I was unable to associate degraded riparian conditions to land uses, the BLM needs to work to determine if these are natural or anthropogenic gradients. If anthropogenic, the BLM needs to decide how they can improve riparian complexity and canopy cover conditions, while implementing their multiple-use mandate.

Challenges to resource monitoring

Biased target populations – An accurate inventory of the type (i.e., intermittent or perennial) of lotic systems is essential to both the efficient implementation of unbiased stream assessment and monitoring programs and the accuracy of population estimates, especially if the program and associated field protocol is addressing one flow permanence type. The NHD is the primary digital representation of aquatic systems for the U.S. and used to define membership in the target population of many stream assessment and monitoring programs (e.g., EPA's NRSA, CA's PSA, AREMP). However, the NHD has been shown to misclassify flow permanence and stream order, and often under represents perennial and intermittent headwater streams (Fritz et al. 2013). I found that 40% of perennial streams classified by the NHD were intermittent at the time of sampling. Such estimates may be high due to an ongoing decadal drought in the study area, but this still represents a major implementation challenge. Stoddard et al. (2005a) found the NHD overestimated hydrologic permanence for 30% of visited sites. In contrast, the PSA found that the NHD underestimated flow permanence in most of CA. I was unable to estimate the later because I did not visit streams classified as intermittent by the NHD.

The problem with inaccuracies or bias in the GIS representation of the target population is that it can increase the cost of implementing assessment and monitoring programs and bias extent estimates (Olsen and Peck 2008). To complicate this problem further it is unrealistic to achieve a static and accurate inventory of flow permanence due to the increasing number of factors that can influence stream flow periodicity, including anthropogenic activities (Milly et al. 2008). One possible solution to dealing with these challenges is to develop methods for desktop screening of sites with remote imagery (e.g., Normalized Difference Vegetation Index, Google Earth, infrared, or several other options) to help determine the likelihood of flow permanence. Such an approach can greatly reduce program costs, as field verification is expensive for remote sites. Another possible and more inclusive solution to dealing with flow permanence inaccuracies is to adapt field protocols to include intermittent streams. Intermittent streams are important resources especially in arid areas, and Mazor et al. (2014) found that intermittent streams can successfully be incorporated into current bioassessment programs. Additionally, including intermittent streams may reduce the ongoing effort needed to refine the NHD flow permanence classification, especially since climate change is likely to exacerbate misclassification issues, as some streams that were once perennial become intermittent (Döll and Schmied 2012, Perry et al. 2012).

The role of benchmarks in adaptive management – Adaptive management is ubiquitously used by federal land management agencies (Williams et al. 2009), however significant impediments remain for effective implementation. Principal among these impediments is the setting of quantitative thresholds, which if exceeded trigger changes in management actions. Fischman and Ruhl (2015) found that the failure to set thresholds was one of the top reasons courts deemed the adaptive management plans of federal agencies to be arbitrary and capricious, as the agency failed to define when alternative management actions would be taken. In attempts to avoid this problem, thresholds for large-scale aquatic monitoring programs are most commonly established using reference conditions, which provide a benchmark for indicator or stressor values expected to occur in the absence of anthropogenic impacts (Stoddard et al. 2006, Herlihy et al. 2008, Hawkins et al. 2010).

The BLM's use of field methods consistent with those of other state, regional, and national monitoring programs allowed me to leverage both reference site networks and analytical tools to set objective thresholds for the condition of a given indicator. However, the use of these reference sites and tools was not without challenges. Principal among these challenges were disparities in reference site networks used to develop threshold values. Ideally, reference sites would be representative of the sampled environmental gradients and selected based on the same criteria for a given region. However, this is not always practical, as different agencies determine reference conditions for disparate applications and for reporting at different spatial scales (e.g., EPA might report on all lands in NV, whereas the BLM only wants to report on the BLM lands in NV). If national (e.g., Hill et al. 2013), regional (Reeves et al. 2003, Olson and Hawkins 2012, 2013), and state-based models (e.g., Mazor et al. 2013, Vander Laan and Hawkins 2014) are all used in one study area for different indicators, and they define reference differently, this could introduce bias in extent estimates among indicators. This use of multiple reference networks also influences the comparability among agencies' stream condition estimates (Miller et al. 2016). However, when trying to be consistent in selecting reference sites it can be very challenging to balance availability of quality reference sites, sample size of reference sites, and the ability to fully capture the environmental variability at the scale at which the stream condition assessments are analyzed (Herlihy et al. 2008).

The use of multiple approaches and models with differing precision and accuracy to set thresholds may bias condition estimates among indicators. Accuracy is a measure of systematic error from the true mean and is influenced by how well the model accounts for natural environmental variability among sites. Although generally more accurate than the regional reference approach, few modeling approaches are completely accurate and thus they can still result in over or under protection of resources depending on the direction of the systematic error (Vander Laan and Hawkins 2014). Precision is a measure of variability around the mean and is influenced by factors such as the predictability of conditions at a site (Vander Laan and Hawkins 2014). As such, an imprecise approach or model may increase the likelihood of under protecting resources because of increases to the upper and lower percentiles of expected values, which are often used as criteria for setting thresholds. This problem is further confounded if model precision is not consistent among indicators. For example, the predictive models I used for excess nutrients had different levels of precision; the model for TN was more precise than the model for TP (Olson and Hawkins 2013). Such differences in model precision among indicators may bias extent estimates and the relative ranking of stressors on the landscape. Models are generally more accurate and precise than the regional reference approach because a modeling approach accounts for the natural environmental variability among sites (Hawkins et al. 2010, Olson and Hawkins 2013, Vander Laan and Hawkins 2014). Moreover, although predictive models do not completely solve the issues discussed above, a predictive model approach has a known level of accuracy and precision, which can be used to balance the likelihood of over versus under protection. In contrast, such information is not available when using the regional reference or best professional judgement approach.

Even with appropriate models and thresholds, the use of empirical monitoring data in landscape-scale adaptive management also requires managers to consider how much of a resource (e.g., what percent of stream km) must exceed a threshold and how confident they need to be in the estimate before changes in management are needed. The CWA requires the restoration of all degraded conditions while acknowledging natural disturbances. Disturbances such as floods and droughts naturally alter macroinvertebrate assemblages and physical conditions such as bank stability or riparian habitats (Resh et al. 1988, Lake 2000, Cardinale et al. 2005). Because some sites may be in non-reference conditions due to these natural disturbances, land management agencies must weigh the allowable extent of a resource in non-reference conditions with management objectives for a given landscape. Furthermore, depending on societal and agency values for a management unit, we need to accept a larger degree of alteration in some areas than other areas. For example, managers are likely to allow less departure from reference in a wilderness area than non-wilderness areas permitted for multiple uses such as livestock grazing and oil and gas development. Regardless of the management objectives, achieving agency goals of sustainable land management requires knowledge of ecosystem resistance and resilience, which can differ among regions, to avoid irreversible degradation and the loss of ecosystem function (Kauffman et al. 1997, Elmqvist et al. 2003, Lake 2013). This latter challenge of allowable degrees of departure is far more vexing a challenge than setting threshold values.

Data and capacity limitations – Through the process of implementing the BLM's AIM strategy and developing a framework for applying the data to management decisions, I encountered numerous challenges and limitations including the measured indicators, inadequate land use and surface disturbance data, and the capacity of BLM resource staff to implement the framework. For example, limited field capacity precluded the deployment of thermistors to quantify seasonal and daily thermal regimes, a variable commonly associated with degraded biological conditions (Daufresne et al. 2007, Krno and Holubec 2009, Chinnayakanahalli et al. 2011). Including this indicator in my biologically relevant stressors RF model could have helped to account for more variability in MMI scores and provided a more complete picture of limiting environmental conditions. However, collecting stream temperature data has some logistical challenges in that it requires at least two visits to each sampled site, first to deploy and second to collect or download temperature loggers. Doing this at every sample reach can be labor intensive and expensive throughout the remote lands managed by the BLM. One solution to this problem is to use predictive models such as the NorWeSt stream temperature model (Isaak et al. 2017) as a first cut to identifying areas of temperature concern and then implement a more intensive monitoring regime for stream temperature in these high priority areas. Riparian complexity is another field method that was limiting. This indicator would be improved by including a native and non-native component, as well as the presence of upland vegetation encroachment. These improvements to the riparian vegetative protocol would help managers better understand the potential drivers of low riparian complexity.

Beyond field data, the geospatial data characterizing land uses and surface disturbances also need to be improved to better interpret and identify land use associations and inform management actions. The GIS layers I used to calculate land uses and surface disturbances were of varying accuracy and resolution which likely confounded land use associations with measured instream conditions. For example, Falcone et al. (2010) found that 42% of the watersheds classified by the USEPA as heavily impacted (with site scale data and aerial imagery) were misclassified when assessed using GIS data alone (e.g., national coverages of agriculture and mining activities). Complicating this further, some of the data I used for this analysis, such as the grazing data, are not stored in a spatial database, adding another level of labor-intensive work for those running analyses.

Lastly, as the BLM embarks on this new endeavor to use consistent field and analytical methods to make data driven management decisions, they will need to ensure there is adequate support, skills, and tools available to help BLM resource staff collect, analyze, and interpret the data. In particular, the BLM needs to ensure additional funds and time availability of resource staff to attend trainings and learn not just how to collect data, but how to use the data to inform management actions. The creation of automated tools to simply plug data into and receive standardized data summary output can greatly increase the efficiency of analyses, but they still requires training on how to interpret the results and use them in management decisions. Additionally, the creation of these tools is labor intensive and requires advanced skills and abilities. To overcome these issues the BLM needs to ensure there are aquatic personnel and scarce skills specialists to train, build analytical tools, and support data use for all resource staff. Additionally, the BLM needs to continue to work with partners and universities to help advance scientific techniques to improve upon the methods I have outlined in this research (e.g., continued development or improvement of predictive models) and develop other tools that would be useful to more efficient and effective land management.

CONCLUSION

The BLM’s AIM-NAMF represents a significant step but rather data towards providing the BLM with the defensible landscape-scale aquatic monitoring data needed to assess the efficacy of management actions at various spatial scales and ensure compliance with federal regulations (e.g., FLPMA, CWA). The most significant challenges facing the effort are not necessarily what to measure or how to measure a given indicator, interpretation. In particular, the setting of meaningful thresholds and allowable degrees of departure from these thresholds to protect the beneficial uses of lotic systems throughout the National System of Public Lands. In addition to needing tools to objectively quantify resource condition and trend, managers must also be able to identify the likely causes of degradation to assist in the adaptive management process. My modeling efforts of land-uses associated with observed stressors produced mixed results and would likely benefit from improved spatial databases of permitted uses throughout the West such as grazing, oil and gas development, and timber harvest. As the BLM continues to implement the AIM-NAMF we need to continue improving field protocols to incorporate intermittent streams, standard tools to quantify lotic condition and trend, and methods to identify the stressors and land uses associated with degraded conditions.

REFERENCES

Allan, J. D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual review of ecology, evolution, and systematics 35:257-284.

Asner, G. P., A. J. Elmore, L. P. Olander, R. E. Martin, and A. T. Harris. 2004. Grazing Systems, Ecosystem Responses, and Global Change. Annual Review of Environment and Resources 29:261-299.

Barquín, J. and R. G. Death. 2004. Patterns of invertebrate diversity in streams and freshwater springs in Northern Spain. Archiv für Hydrobiologie 161:329-349.

Beschta, R. L., D. L. Donahue, D. A. DellaSala, J. J. Rhodes, J. R. Karr, M. H. O'Brien, T. L. Fleischner, and C. Deacon Williams. 2012. Adapting to Climate Change on Western Public Lands: Addressing the Ecological Effects of Domestic, Wild, and Feral Ungulates. Environmental Management 51:1-18.

BLM. 2015. AIM National Aquatic Monitoring Framework: Introducing the Framework and Indicators for Lotic Systems. Technical Reference 1735-1. U.S. Department of the Interior, Bureau of Land Management, National Operations Center, Denver, CO.

Bonada, N., N. Prat, V. H. Resh, and B. Statzner. 2006. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. Annual Review of Entomology 51:495-523.

Bourassa, N. and A. Cattaneo. 1998. Control of periphyton biomass in Laurentian streams (Québec). Journal of the North American Benthological Society 17:420-429.

Breiman, L. 2001. Random forests. Machine learning 45:5-32.

Burton, T. A., S. J. Smith, and E. R. Cowley. 2011. Multiple Indicator Monitoring (MIM) of Stream Channels and Streamside Vegetation. US Department of the interior, Bureau of Land Management, Technical Reference:1737-1723.

Cao, Y. and C. P. Hawkins. 2011. The comparability of bioassessments: a review of conceptual and methodological issues. Journal of the North American Benthological Society 30:680-701.

Cardinale, B. J., M. A. Palmer, A. R. Ives, and S. Brooks. 2005. Diversity-productivity relationships in streams vary as a function of the natural disturbance regime. Ecology 86:716-726.

Carlisle, D., J. Falcone, and M. Meador. 2009. Predicting the biological condition of streams: use of geospatial indicators of natural and anthropogenic characteristics of watersheds. Environmental Monitoring and Assessment 151:143-160.

Chambers, P. A., D. J. McGoldrick, R. B. Brua, C. Vis, J. M. Culp, and G. A. Benoy. 2012. Development of environmental thresholds for nitrogen and phosphorus in streams. Journal of Environmental Quality 41:7-20.

Chinnayakanahalli, K., C. Kroeber, and R. Hill. 2006. The multiwatershed delineation tool: GIS software in support of regional watershed analyses. Utah State University, Logan.

Chinnayakanahalli, K. J., C. P. Hawkins, D. G. Tarboton, and R. A. Hill. 2011. Natural flow regime, temperature and the composition and richness of invertebrate assemblages in streams of the western United States. Freshwater Biology 56:1248-1265.

Cross, W. F., B. R. Johnson, J. B. Wallace, and A. D. Rosemond. 2005. Contrasting response of stream detritivores to long-term nutrient enrichment. Limnology and Oceanography 50:1730-1739.

Culp, J., K. Cash, and F. Wrona. 2000. Cumulative effects assessment for the Northern River Basins Study. Journal of Aquatic Ecosystem Stress and Recovery 8:87-94.

Cutler, D. R., T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

Daufresne, M., P. Bady, and J.-F. Fruget. 2007. Impacts of global changes and extreme hydroclimatic events on macroinvertebrate community structures in the French Rhône River. Oecologia 151:544.

del Rosario, R. B., E. A. Betts, and V. H. Resh. 2002. Cow manure in headwater streams: tracing aquatic insect responses to organic enrichment. Journal of the North American Benthological Society 21:278-289.

Delong, M. D. and M. A. Brusven. 1994. Allochthonous input of organic matter from different riparian habitats of an agriculturally impacted stream. Environmental Management 18:59-71.

Delong, M. D. and M. A. Brusven. 1998. Macroinvertebrate community structure along the longitudinal gradient of an agriculturally impacted stream. Environmental Management 22:445-457.

Dodds, W. K. 2007. Trophic state, eutrophication and nutrient criteria in streams. Trends in ecology & evolution 22:669-676.

Dodds, W. K. and E. B. Welch. 2000. Establishing nutrient criteria in streams. Journal of the North American Benthological Society 19:186-196.

Döll, P. and H. M. Schmied. 2012. How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. Environmental Research Letters 7:014037.

Elmqvist, T., C. Folke, M. Nyström, G. Peterson, J. Bengtsson, B. Walker, and J. Norberg. 2003. Response diversity, ecosystem change, and resilience. Frontiers in Ecology and the Environment 1:488-494.

Falcone, J. A., D. M. Carlisle, and L. C. Weber. 2010. Quantifying human disturbance in watersheds: Variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. Ecological Indicators 10:264-273.

Fischman, R. L. and J. Ruhl. 2015. Judging adaptive management practices of US agencies. Conservation Biology 30:268-275.

Fleischner, T. L. 1994. Ecological costs of livestock grazing in western North America. Conservation Biology 8:629-644.

Folt, C. L., C. Y. Chen, M. V. Moore, and J. Burnaford. 1999. Synergism and antagonism among multiple stressors. Limnology and Oceanography 44:864-877.

Fritz, K. M., E. Hagenbuch, E. D'Amico, M. Reif, P. J. Wigington, S. G. Leibowitz, R. L. Comeleo, J. L. Ebersole, and T. L. Nadeau. 2013. Comparing the extent and permanence of headwater streams from two field surveys to values from hydrographic databases and maps. Journal of the American Water Resources Association 49:867-882.

Gomi, T., R. C. Sidle, M. D. Bryant, and R. D. Woodsmith. 2001. The characteristics of woody debris and sediment distribution in headwater streams, southeastern Alaska. Canadian Journal of Forest Research 31:1386-1399.

Greenwood, J., A. Rosemond, J. B. Wallace, W. Cross, and H. Weyers. 2007. Nutrients stimulate leaf breakdown rates and detritivore biomass: bottom-up effects via heterotrophic pathways. Oecologia 151:637-649.

Hart, D. D. and C. T. Robinson. 1990. Resource limitation in a stream community: phosphorus enrichment effects on periphyton and grazers. Ecology 71:1494-1502.

Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessments. Ecological applications 16:1277-1294.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological applications 10:1456-1477.

Hawkins, C. P., J. R. Olson, and R. A. Hill. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society 29:312-343.

Hawkins, C. P., S. G. Paulsen, J. V. Sickle, and L. L. Yuan. 2008. Regional assessments of stream ecological condition: scientific challenges associated with the USA's national Wadeable Stream Assessment. Journal of the North American Benthological Society 27:805-807.

Hayashi, M., T. Vogt, L. Mächler, and M. Schirmer. 2012. Diurnal fluctuations of electrical conductivity in a pre-alpine river: Effects of photosynthesis and groundwater exchange. Journal of Hydrology 450:93-104.

Herbst, D. B., M. T. Bogan, S. K. Roll, and H. D. Safford. 2012. Effects of livestock exclusion on in-stream habitat and benthic invertebrate assemblages in montane streams. Freshwater Biology 57:204-217.

Herlihy, A. T., S. G. Paulsen, J. V. Sickle, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. Journal of the North American Benthological Society 27:860-877.

Hill, R. A., C. P. Hawkins, and D. M. Carlisle. 2013. Predicting thermal reference conditions for USA streams and rivers. Freshwater Science 32:39-55.

Hughes, S. J., J. M. Santos, M. T. Ferreira, R. Caraça, and A. M. Mendes. 2009. Ecological assessment of an intermittent Mediterranean river using community structure and function: Evaluating the role of different organism groups. Freshwater Biology 54:2383-2400.

Ice, G. and D. Binkley. 2003. Forest streamwater concentrations of nitrogen and phosphorus: A comparison with EPA's proposed water quality criteria. Journal of Forestry 101:21-28.

Isaak, D. J., S. J. Wenger, E. E. Peterson, J. M. Ver Hoef, D. E. Nagel, C. H. Luce, S. W. Hostetler, J. B. Dunham, B. B. Roper, S. P. Wollrab, G. L. Chandler, D. L. Horan, and S. Parkes-Payne. 2017. The NorWeST summer stream temperature model and scenarios for the western U.S.: A crowd-sourced database and new geospatial tools foster a user community and predict broad climate warming of rivers and streams. Water Resources Research 53:9181-9205.

ITFM. 1995. The strategy for improving water-quality monitoring in the United States. Final Report 95-742. United States Geological Survey, Office of Water Data Coordination, Reston, VA.

Kauffman, J. B., R. L. Beschta, N. Otting, and D. Lytjen. 1997. An ecological perspective of riparian and stream restoration in the western United States. Fisheries 22:12-24.

Kershner, J. L., E. K. Archer, M. Coles-Ritchie, E. R. Cowley, R. Henderson, K. Kratz, C. M. Quimby, D. L. Turner, L. C. Ulmer, and M. R. Vinson. 2004. Guide to effective monitoring of aquatic and riparian resources. U.S. Dept. of Agriculture, Forest Service, Rocky Mountain Research Station.

Kincaid, T., A. Olsen, D. Stevens, C. Platt, D. White, and R. Remington. 2013. spsurvey: Spatial Survey Design and Analysis. R package version 2.6.

Krno, I. j. and M. Holubec. 2009. Effects of land use on stonefly bioassessment metrics. Aquatic Insects 31:377-389.

Lake, P. S. 2000. Disturbance, patchiness, and diversity in streams. Journal of the North American Benthological Society 19:573-592.

Lake, P. S. 2013. Resistance, resilience and restoration. Ecological Management & Restoration 14:20-24.

Liaw, A. and M. Wiener. 2002. Classification and regression by randomForest. R news 2:18-22.

Marzin, A., V. Archaimbault, J. Belliard, C. Chauvin, F. Delmas, and D. Pont. 2012. Ecological assessment of running waters: Do macrophytes, macroinvertebrates, diatoms and fish show similar responses to human pressures? Ecological Indicators 23:56-65.

Matthaei, C. D., J. J. Piggott, and C. R. Townsend. 2010. Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. Journal of Applied Ecology 47:639-649.

Mazor, R. D., A. Rehn, P. R. Ode, M. Engeln, and S. K. 2013 *Draft*. Maximizing the applicability of bioassessment scoring tools in environmentally complex regions.

Mazor, R. D., E. D. Stein, P. R. Ode, and K. Schiff. 2014. Integrating intermittent streams into watershed assessments: applicability of an index of biotic integrity. Freshwater Science 33:459-474.

McDonald, T. 2012. Spatial sampling designs for long-term ecological monitoring.*in* R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, editors. Design and Analysis of Long-term Ecological Monitoring Studies.

Miller, S., P. Eldred, A. Muldoon, K. Anlauf-Dunn, C. Stein, S. Hubler, L. Merrick, N. Haxton, C. Larson, A. Rehn, P. Ode, and J. Vander Laan. 2016. A Large-Scale, Multiagency Approach to Defining a Reference Network for Pacific Northwest Streams. Environmental Management 58:1091-1104.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer. 2008. Stationarity Is Dead: Whither Water Management? Science 319:573-574.

Miltner, R. J. and E. T. Rankin. 1998. Primary nutrients and the biotic integrity of rivers and streams. Freshwater Biology 40:145-158.

Nadeau, T. L. and M. C. Rains. 2007. Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy1. JAWRA Journal of the American Water Resources Association 43:118-133.

Naiman, R. J. and H. Décamps. 1997. The ecology of interfaces: riparian zones. Annual review of Ecology and Systematics 28:621-658.

Niyogi, D. K., K. S. Simon, and C. R. Townsend. 2003. Breakdown of tussock grass in streams along a gradient of agricultural development in New Zealand. Freshwater Biology 48:1698-1708.

Ode, P. R., T. M. Kincaid, T. Fleming, and A. C. Rehn. 2009. Ecological condition assessments of California's perennial wadeable streams: highlights from the Surface Water Ambient Monitoring Program's Perennial Streams Assessment (PSA) (2000-2007). A collaboration between the State Water Resources Control Board's Non-Point Source Pollution Control Program (NPS Program), Surface Water Ambient Monitoring Program (SWAMP), California Department of Fish and Game Aquatic Bioassessment Laboratory, and the U.S. Environmental Protection Agency.

Olsen, A. R., T. M. Kincaid, and Q. Payton. 2012. Spatially balanced survey designs for natural resources.*in* R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, editors. Design and Analysis of Long-term Ecological Monitoring Studies.

Olsen, A. R. and D. V. Peck. 2008. Survey design and extent estimates for the Wadeable Streams Assessment. Journal of the North American Benthological Society 27:822-836.

Olson, J. R. and C. P. Hawkins. 2012. Predicting natural base-flow stream water chemistry in the western United States. Water Resources Research 48:W02504.

Olson, J. R. and C. P. Hawkins. 2013. Developing site-specific nutrient criteria from empirical models. Freshwater Science 32:719-740.

Omernik, J. M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American Geographers 77:118-125.

Ostermiller, J. D. and C. P. Hawkins. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. Journal of the North American Benthological Society 23:363-382.

Pang, H., A. Lin, M. Holford, B. E. Enerson, B. Lu, M. P. Lawton, E. Floyd, and H. Zhao. 2006. Pathway analysis using random forests classification and regression. Bioinformatics 22:2028-2036.

Passy, S. I., R. W. Bode, D. M. Carlson, and M. A. Novak. 2004. Comparative Environmental Assessment in the Studies of Benthic Diatom, Macroinvertebrate, and Fish Communities. International Review of Hydrobiology 89:121-138.

Paulsen, S. G., R. M. Hughes, and D. P. Larsen. 1998. Critical elements in describing and understanding our nation's aquatic resources Journal of the American Water Resources Association 34:995-1005.

Paulsen, S. G., A. Mayio, D. V. Peck, J. L. Stoddard, E. Tarquinio, S. M. Holdsworth, J. V. Sickle, L. L. Yuan, C. P. Hawkins, A. T. Herlihy, P. R. Kaufmann, M. T. Barbour, D. P. Larsen, and A. R. Olsen. 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. Journal of the North American Benthological Society 27:812-821.

Perrin, C. and J. Richardson. 1997. N and P limitation of benthos abundance in the Nechako River. British Columbia. Canadian.

Perry, L. G., D. C. Andersen, L. V. Reynolds, S. M. Nelson, and P. B. Shafroth. 2012. Vulnerability of riparian ecosystems to elevated CO2 and climate change in arid and semiarid western North America. Global Change Biology 18:821-842.

Peters, J., B. D. Baets, N. E. C. Verhoest, R. Samson, S. Degroeve, P. D. Becker, and W. Huybrechts. 2007. Random forests as a tool for ecohydrological distribution modelling. Ecological Modelling 207:304-318.

Prasad, A., L. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181-199.

Prichard, D., J. Anderson, C. Correll, J. Fogg, K. Gebhardt, R. Krapf, S. Leonard, B. Mitchell, and J. Staats. 1998. Riparian area management: A user guide to assessing proper functioning condition and the supporting science for lotic areas TR 1737-15. US Dept. of Interior-Bureau of Land Management. US Department of Agriculture-Forest Service and USDA Natural Resources Conservation Service.

R Core Team. 2012. R: A Language and Environment for Statistical Computing, Vienna, Austria.

Rader, R. B., M. J. Keleher, E. Billman, and R. Larsen. 2012. History, rather than contemporary processes, determines variation in macroinvertebrate diversity in artesian springs: the expansion hypothesis. Freshwater Biology 57:2475-2486.

Reeves, G. H., D. B. Hohler, D. P. Larsen, D. E. Busch, K. Kratz, K. Reynolds, K. F. Stein, T. Atzet, P. Hays, and M. Tehan. 2003. Aquatic and riparian effectiveness monitoring plan for the Northwest Forest Plan. USDA Forest Service, Pacific Northwest Research Station.

Resh, V. H., A. V. Brown, A. P. Covich, M. E. Gurtz, H. W. Li, G. W. Minshall, S. R. Reice, A. L. Sheldon, J. B. Wallace, and R. C. Wissmar. 1988. The role of disturbance in stream ecology. Journal of the North American Benthological Society 7:433-455.

Rios, S. and R. Bailey. 2006. Relationship between Riparian Vegetation and Stream Benthic Communities at Three Spatial Scales. Hydrobiologia 553:153-160.

Robinson, C. T. and M. O. Gessner. 2000. Nutrient addition accelerates leaf breakdown in an alpine springbrook. Oecologia 122:258-263.

Sada, D. W., E. Fleishman, and D. D. Murphy. 2005. Associations among spring-dependent aquatic assemblages and environmental and land use gradients in a Mojave Desert mountain range. Diversity and Distributions 11:91-99.

Schreuder, H., T. Gregoire, and J. Weyer. 2001. For what applications can probability and non-probability sampling be used? Environmental Monitoring and Assessment 66:281-291.

Sidle, R. C., Y. Tsuboyama, S. Noguchi, I. Hosoda, M. Fujieda, and T. Shimizu. 2000. Stormflow generation in steep forested headwaters: a linked hydrogeomorphic paradigm. Hydrological Processes 14:369-385.

Stevens, D. L. and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99:262-278.

Stoddard, J. L., D. V. Peck, A. R. Olsen, D. P. Larsen, J. Van Sickle, C. P. Hawkins, R. M. Hughes, T. R. Whittier, G. Lomnicky, and A. T. Herlihy. 2005a. Environmental Monitoring and Assessment Program (EMAP): western streams and rivers statistical summary. Washington, DC: US Environmental Protection Agency, Office of Research and Development.

Stoddard, J. L., D. V. Peck, S. G. Paulsen, J. Van Sickle, C. P. Hawkins, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. P. Larsen, G. Lomnicky, A. R. Olsen, S. A. Peterson, P. L. Ringold, and T. R. Whittier. 2005b. An ecological assessment of western streams and rivers. EPA 620/R-05/005, U.S. Environmental Protection Agency, Washington, DC.

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878-891.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological applications 16:1267-1276.

Toevs, G. R., J. W. Karl, J. J. Taylor, C. S. Spurrier, M. S. Karl, M. R. Bobo, and J. E. Herrick. 2011. Consistent indicators and methods and a scalable sample design to meet assessment, inventory, and monitoring information needs across scales. Rangelands 33:14-20.

Townsend, C. R., S. S. Uhlmann, and C. D. Matthaei. 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45:1810-1819.

USEPA. 2009. National rivers and streams assessment: field operations manual. Environmental Protection Agency, Washington, DC.

USGS. 2012. National Hydrography Dataset Plus - NHDPlus.

Van Sickle, J. and S. G. Paulsen. 2008. Assessing the attributable risks, relative risks, and regional extents of aquatic stressors. Journal of the North American Benthological Society 27:920-931.

Vander Laan, J. J. and C. P. Hawkins. 2014. Enhancing the performance and interpretation of freshwater biological indices: An application in arid zone streams. Ecological Indicators 36:470-482.

Vander Laan, J. J., C. P. Hawkins, J. R. Olson, and R. A. Hill. 2013. Linking land use, in-stream stressors, and biological condition to infer causes of regional ecological impairment in streams. Freshwater Science 32:801-820.

Wang, L., D. M. Robertson, and P. J. Garrison. 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: implication to nutrient criteria development. Environmental Management 39:194-212.

Weigel, B. and D. Robertson. 2007. Identifying biotic integrity and water chemistry relations in nonwadeable rivers of Wisconsin: toward the development of nutrient criteria. Environmental Management 40:691-708.

Williams, B. K., R. C. Szaro, and C. D. Shapiro. 2009. Adaptive management. Technical Guide. The US Department of the Interior 172.

Yuan, L. L. 2010. Estimating the effects of excess nutrients on stream invertebrates from observational data. Ecological applications 20:110-125.

APPENDICES

APPENDIX A

Table A-1. Natural and anthropogenic variables used as variables in NMDS ordinations, empirical models, and random forest models. CA= California biological index, NV OE= Nevada observed to expected index, NV MMI=Nevada Multimetric Index, TN= Total Nitrogen model, TP=Total Phosphorus model, EC=Electrical Conductivity (specific conductance) model, RF Stress= Random Forest biologic condition to stressors model, and RFLU= Random Forest stressor to land use model. CA variables also used to create CA NMDS ordinations, NV OE + NV MMI used to create NV NMDS ordinations.

Metric	Model	Description
A_Prop_3YrPrPr	RFLU	Active animal unit months (AUM) divided by Permitted AUMs for the 3 years prior to the year of sampling for each allotment, then summed for all allotments overlapping sampled watershed. Data acquired from personal communication with the BLM field offices.
A_Prop_YrPr	RFLU	Active animal unit months (AUM) divided by Permitted AUMs for the year prior to the year of sampling for each allotment, then summed for all allotments overlapping sampled watershed. Data acquired from personal communication with the BLM field offices.
A_SpDensity800m	RFLU	Density of springs in an 800 meter radius of sample point within the watershed calculated with the National Hydrography Dataset.
AG_WS	RFLU	The percent agriculture land coverage in the watershed calculated from the National Land Cover Dataset.
alru_dom	TN, RF Stress	Percent are of watershed with Alnus rubra dominated cover from the National Land Cover Dataset.
ArtPathDens	RFLU	The density of artificial paths in the watershed calculated by total length of artificial path divided by area. Path lines from the National Hydrography dataset.
AtmCa	TP, EC	Atmospheric Calcium at the sample point.
AtmMg	EC	Atmospheric Magnesium at the sample point.
AtmNa	TN	Atmospheric Na deposition at the sample point.
AtmNO3	TN	Atmospheric NO3 deposition at the sample point.
AtmSO4	TP, EC	Atmospheric SO4 at the sample point.
AWC_soil	TP	Watershed mean of the high values of available water capacity (fraction) of soils from the State Soil Geographic (STATSGO) Database.

Metric	Model	Description
BDH_AVE	TN, EC, CA	Watershed mean of the high values of soil bulk density (grams/cm3) of soils from the State Soil Geographic (STATSGO) Database.
BFI_WS	NV MMI	Mean of all base flow index pixel values within the watershed. Estimates the percent of stream flow that is composed of ground water relative to event flow. Calculated from USGS generated 1-km resolution grid of base flows derived by interpolating calculated base flows at 19,000 USGS stream flow gauging stations distributed across the conterminous USA.
CaO_Mean	TP, EC, CA	Mean of all cells within the watershed, where cells represent the percent of the underlying bedrock composed of calcium oxide (CaO). Percentages are the average percent CaO for all lithologies within a cell, weighted by lithology prevalence. Lithologies and their prevalence were derived from the USGS Preliminary Integrated Geologic Map of the United States.
DAMden_WS	RFLU	Density of dams in the watershed.
DAMvol_Stand_WS	RFLU	Volume of dams in the watershed, standardized by watershed size.
Density_RdCross	RFLU	Density of roads crossing streams in the watershed. Flow lines from National Hydrography dataset.
DOY	TN	Julian day of year sample collected.
ELEV_RANGE	CA, RF Stress, RFLU	The difference in elevation from the highest point in the watershed and the lowest point in the watershed.
ELVcv_PT	NV MMI	Coefficient of variation of elevations within a radius of 5 digital elevation model cells (30 m × 30 m resolution) of the sample site.
ELVmax_WS	NV MMI, RFLU	Maximum watershed elevation in meters.
ELVmean_WS	NV MMI, NV OE, RFLU	Mean watershed elevation in meters.
ELVmin_WS	NV MMI	Minimum watershed elevation in meters.
ER13	TP	Presence or absence of Gila Mountains Ecoregion.
Evergr_ave	TN	Percent area of watershed with evergreen vegetation dominated cover from the National Land Cover Dataset.
EVI_AveAve	TN,TP	Mean Enhanced Vegetation Index.

Metric	Model	Description
EVI_MaxAve	EC	Mean of maximum Enhanced Vegetation Index.
GW_P_Sp_Mx	TN	Maximum Ground Water Index in the watershed.
HYDR_WS	NV MMI, NV OE, RF Stress, RFLU	Mean of all point values within the watershed. Point values were calculated with a GIS raster calculated as (MINxi)/(MAXxi), where xi= mean monthly discharge for month i for the period of record and xi≥ 12 months of record. Values were calculated for each of 9941 USGS gauging stations in the western USA and values for unmeasured locations were interpolated using inverse-distance-squared weighting of the 12 closest gauging stations within 100 km. Each interpolated value represents a 4 km × 4 km cell.
IntDensC	RF Stress, RFLU	The density of intermittent flow lines in the watershed calculated by total length of intermittent streams divided by area. Flow lines from the National Hydrography dataset.
KFCT_AVE	EC, CA, TP, RF Stress, RFLU	Watershed mean of the soil erodibility factor (no units) of soils from the State Soil Geographic (STATSGO) Database.
LPREM_mean	EC, CA	Log of mean hydraulic conductivity of the watershed.
LST32AVE	EC	30 year mean of last freeze day in the watershed from PRISM data.
MAXWD_WS	EC	30 year mean of max number of wet days in the watershed from PRISM data.
MEANP_WS	EC, RF Stress, RFLU	Watershed mean precipitation in millimeters.
MgO_Mean	EC, CA	Mean of all cells within the watershed, where cells represent the percent of the underlying bedrock composed of magnesium oxide (MgO). Percentages are the average percent MgO for all lithologies within a cell, weighted by lithology prevalence. Lithologies and their prevalence were derived from the USGS Preliminary Integrated Geologic Map of the United States.
MINEden_WS	RFLU	Density of mines in the watershed.
MINP_WS	EC	30 year mean of the minimum precipitation values within the watershed from PRISM data.
N_MEAN	CA	Average total nitrogen within the watershed.
New_Lat	CA	Latitude in decimal degrees.

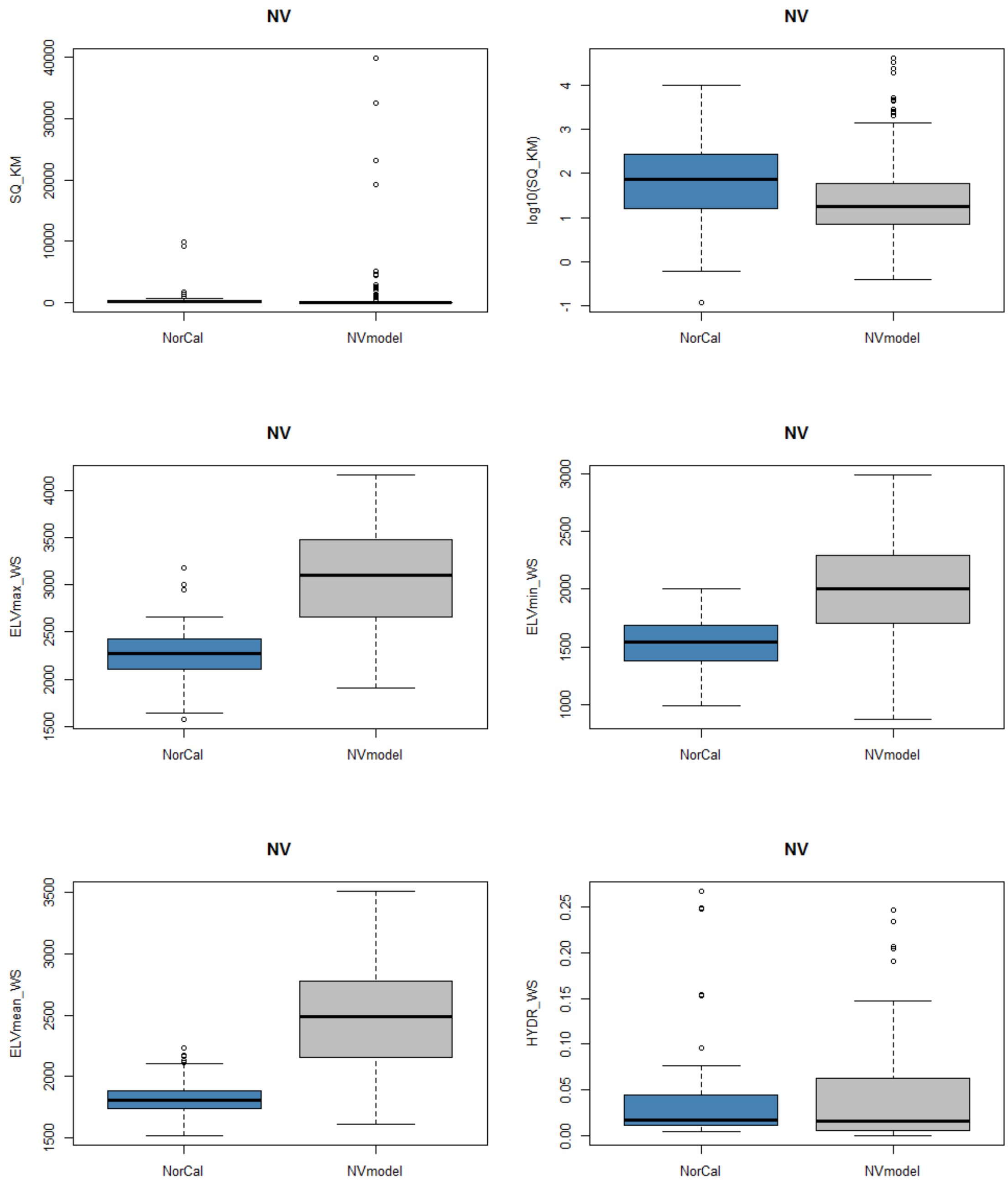
Metric	Model	Description
New_Long	CA	Longitude in decimal degrees.
P_MEAN	CA	Average total phosphorus within the watershed.
Pct_Alfi	TP	Percent alfi soils in the watershed.
PCT_SEDIM	CA, RFLU	Percentage of watershed that is sedimentary geology type
PctXclsr	RFLU	Percent livestock exclosures in the watershed. Data obtained from personal communication with BLM field office personnel.
Percent_HMA	RFLU	Percent Horse and Burro Management Herds in the watershed. Data obtained from personal communication with BLM field office personnel.
PerDensC	RF Stress, RFLU	The density of perennial flow lines in the watershed calculated by total length of perennial streams divided by area. Flow lines from the National Hydrography dataset.
Pmax_PT	NV MMI	30 year average maximum precipitation at the sample point calculated from PRISM data.
Pmax_WS	NV MMI, NV OE	30 year average maximum precipitation of the watershed calculated from PRISM data.
Pmin_WS	NV MMI	30 year average minimum precipitation of the watershed calculated from PRISM data.
PPT_00_09	CA	Average precipitation at the sample point.
PPT_2MoAvg	TN	Mean of the previous month's precipitation and the current month's precipitation calculated from PRISM data.
PPT_ACCUM	TP	Mean of previous year's precipitation sum (May-April) calculated from PRISM data.
PrdCond	NV MMI, NV OE	Expected specific conductance at sampling point (Olson and Hawkins 2012)
PRMH_AVE	EC, CA, RF Stress, RFLU	Watershed mean of the high values of permeability (inches/hour) of soils from the State Soil Geographic (STATSGO) Database
PT_Tmin	TP	Minimum temperature of 30 year mean at the sample point calculated with PRSIM data.

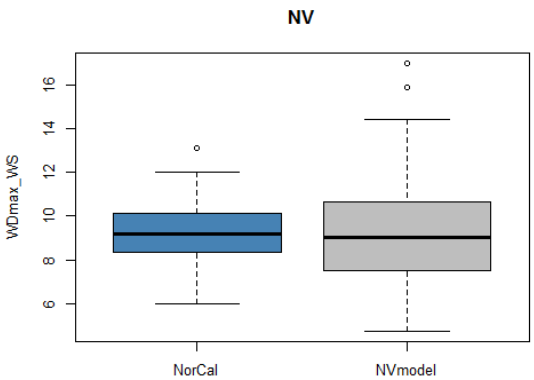
Metric	Model	Description
RdDensC	RFLU	The density of roads in the watershed calculated by total length of roads divided by area.
RH_WS	TP	Mean of all relative humidity values within the watershed from 30 year PRISM data.
S_Mean	EC, CA	Mean of underlying bedrock composed of sulfur(S) within the watershed from the USGS Preliminary Integrated Geologic Map of the United States.
SITE_ELEV	CA, RF Stress, RFLU	Elevation at the sample point.
Slope_WS	NV MMI, NV OE, TP, RF Stress, RFLU	Watershed slope measured as the (ELVmax WS − ELVmin WS)/Maximum flow length.
SOC	TP	Soil organic carbon for the watershed.
SpDensity300m	RFLU	Density of springs in a 300 meter radius of sample point within the watershed calculated with the National Hydrography Dataset.
SpNum300m	RF Stress	Number of springs in a 300 meter radius of sample point within the watershed calculated with the National Hydrography Dataset.
SpNum800m	RF Stress	Number of springs in an 800 meter radius of sample point within the watershed calculated with the National Hydrography Dataset.
SprgDensity_WS	RFLU	Density of springs in the watershed calculated with the National Hydrography Dataset.
SprgNum_WS	RF Stress	Number of springs in the watershed calculated with the National Hydrography Dataset.
SQ_KM	NV MMI, NV OE, CA, RF Stress, RFLU	Watershed area in square km
StmOrd	RFLU	Strahler stream order calculated from the National Hydrologic Dataset Plus

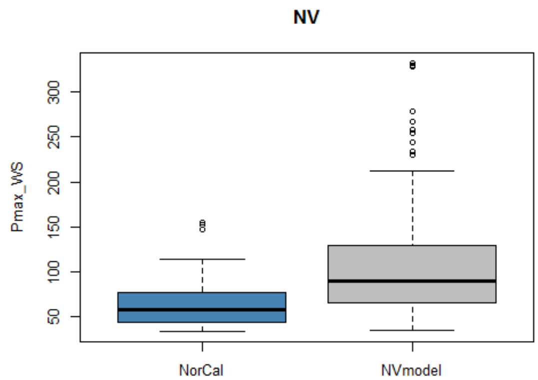
Metric	Model	Description
StreamDens	RF Stress, RFLU	The density of flow lines in the watershed calculated by total length of streams divided by area. Flow lines from the National Hydrography dataset
SumAve_P	EC, CA, RF Stress, RFLU	Mean summer precipitation for the watershed.
TEMP_00_09	CA	Average temperature at the sample point.
Tmax_PT	NV MMI	Maximum temperature at the sample point calculated from 30 year mean of maximum temperatures from PRISM data.
Tmax_WS	NV MMI, NV OE, EC, RF Stress, RFLU	Average maximum temperature of the watershed calculated from 30 year mean of maximum temperatures from PRISM data.
TMIN_WS	TN, RF Stress, RFLU	Average minimum temperature of the watershed calculated from 30 year mean of minimum temperatures from PRISM data.
TP_Mean	TP	Mean of total phosphorus cells in the watershed.
UCS_Mean	EC, RF Stress, RFLU	Mean of the Mean uniaxial compressive strength of the watershed from the USGS Preliminary Integrated Geologic Map of the United States.
URBAN_WS	RFLU	The percent urban coverage in the watershed calculated from the National Land Cover Dataset.
Vol_ave	TP, RFLU	Percentage of watershed that is volcanic geology type.
Wb_mx_area	TP	Area of the largest water body within watershed.
WDmax_WS	NV MMI, TN, EC	Mean of maximum number of wet days in the watershed from 30 year PRISM data.

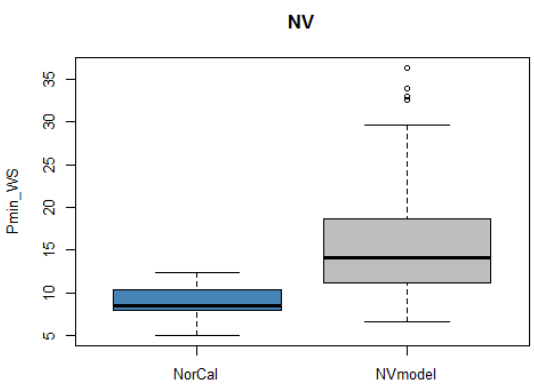
APPENDIX B

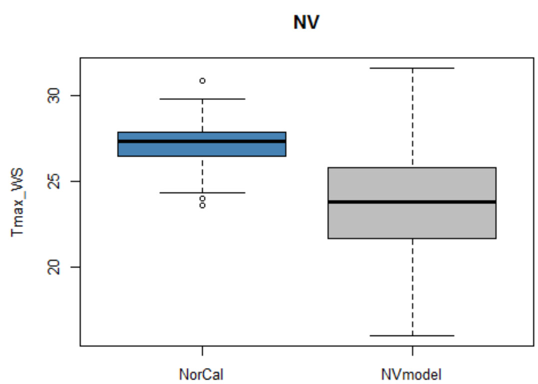
Fig. B-1. Boxplots of watershed and point data used to help determine if there were specific test sites to which I should not apply the NV MMI. Reference sites used to develop the NV MMI are labeled as NV model and represented by grey boxes. Test sites are labeled as NorCal and represented by blue boxes. Variables along the Y-axis are described in Table A-1.

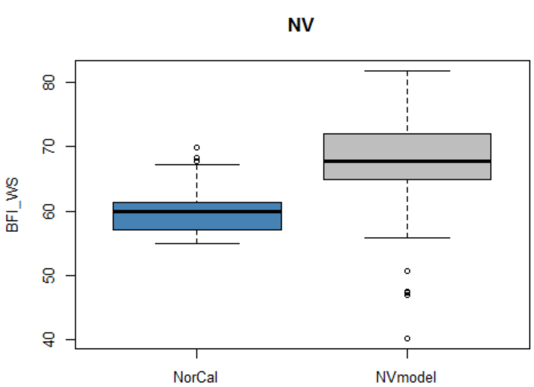


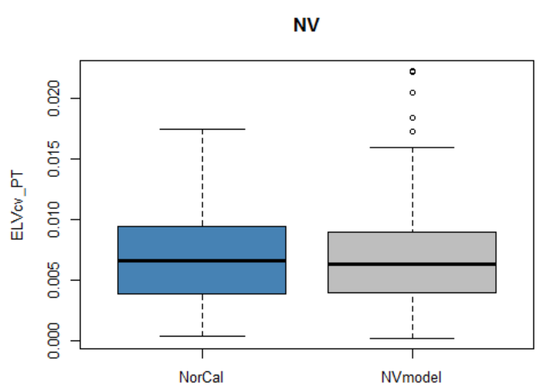

NV
WDmax_WS
6
8
10
12
14
16
NorCal
NVmodel


NV
Pmax_WS
50
100
150
200
250
300
NorCal
NVmodel


NV
Pmin_WS
5
10
15
20
25
30
35
NorCal
NVmodel


NV
Tmax_WS
20
25
30
NorCal
NVmodel


NV
BFI_WS
40
50
60
70
80
NorCal
NVmodel


NV
ELVcv_PT
0.000
0.005
0.010
0.015
0.020
NorCal
NVmodel

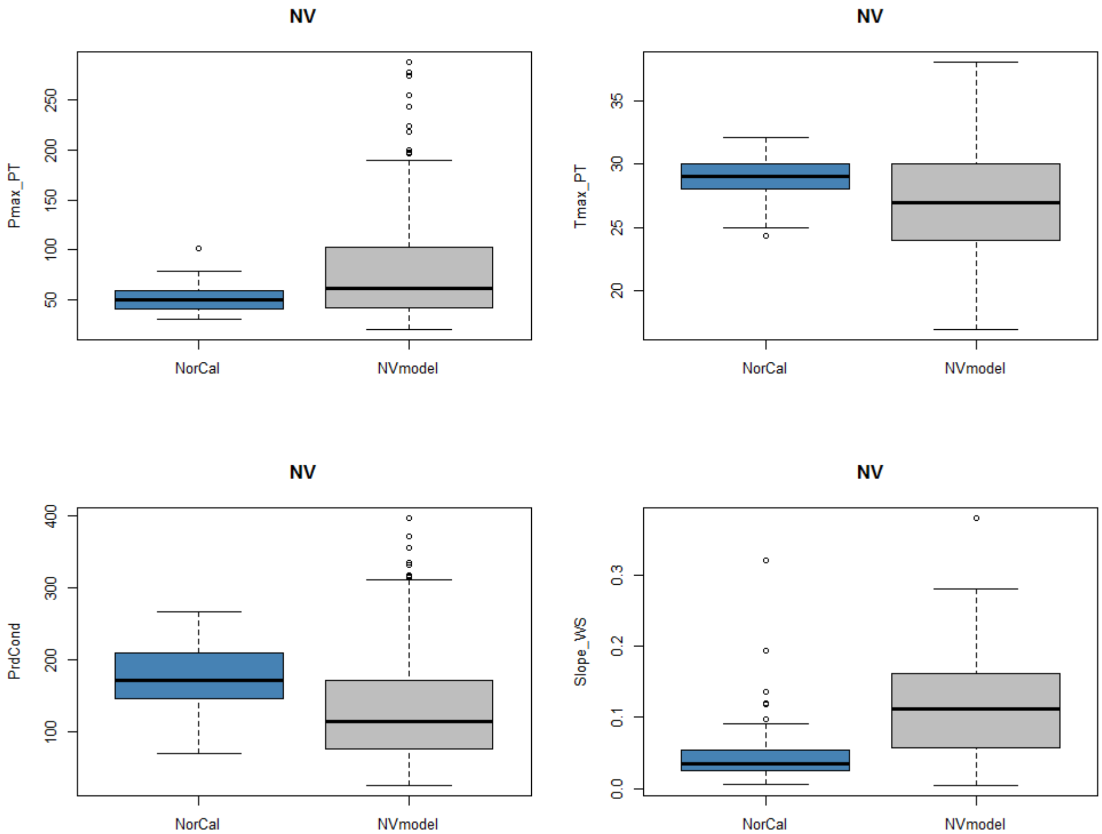


Fig. B-2. Scatter plots of test sites (blue points) and NV MMI reference sites (blue points) to help identify potential test site outliers to which I should not apply the NV MMI. The three circled points were excluded based on results from NMDS ordinations, boxplots (Fig. B-1), and these scatterplots.